Structured Output Prediction and Binary Code Learning in Computer Vision

by

Guosheng Lin

A thesis submitted in fulfillment for the
degree of Doctor of Philosophy

in the
Faculty of Engineering, Computer and Mathematical Sciences
School of Computer Science

December 2014

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the Universitys digital research repository, the Library catalogue and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed: ________________________

Date: ________________________

Acknowledgements

Above all, I would like to express my profound gratitude to my principal supervisor Prof. Chunhua Shen for his continued guidance and strong support throughout my time as his student. He taught me not only fundamental and advanced knowledge of machine learning, but also scientific thinking and writing skills. He gave me insightful guidance and creative suggestions on my research problems. He made himself available to me at any time for discussing research problems. This thesis would not have been possible without his guidance.

I would like to thank my co-supervisor Prof. David Suter. He encouraged me working on my research interests, helped me to develop scientific thinking, inspired me thinking outside the box, and improved my presentation skills. I would also like to thank my co-supervisor Dr.Tat-Jun Chin. His questions raised in meetings helped me deliver clear and convincing presentations of my ideas.

I would like to thank Prof. Anton van den Hengel for his great support and invaluable collaboration. He gave me important suggestions on presentation, and offer great help on paper writing and revision. I am also grateful to Dr. Qinfeng (Javen) Shi. His insightful viewpoints and suggestions helped me improve my research methodologies and paper writing.

Many thanks go to ACVT researchers (current or previous) for their kindness and beneficial discussions: Dr. Xi Li, Dr. Sakrapee (Paul) Paisitkriangkrai, Dr. Peng Wang, Dr. Lingqiao Liu and Dr. Jin Yu. I would also like to thank all of my lab mates, especially Quoc-Huy Tran, Zhenhua Wang, Rui Yao, Fumin Shen and Zhen Zhang.

Finally, my special appreciation is dedicated to my parents, brother, sister for their everlasting support, and to my girlfriend with whom I share my success and frustration.

Publications

This thesis is based on the content of the following peer-reviewed conference and journal papers:

1. Guosheng Lin, Chunhua Shen, Qinfeng Shi, Anton van den Hengel and David Suter; *"Fast Supervised Hashing with Decision Trees for High-Dimensional Data"*; In proceeding of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.

2. Guosheng Lin, Chunhua Shen, Jianxin Wu;
"Optimizing Ranking Measures for Compact Binary Code Learning"; In proceeding of European Conference on Computer Vision (ECCV), 2014.

3. Chunhua Shen, Guosheng Lin, Anton van den Hengel;
"StructBoost: Boosting Methods For Predicting Structured Output Variables"; IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2014.

4. Guosheng Lin, Chunhua Shen, David Suter and Anton van den Hengel;
"A General Two-Step Approach to Learning-Based Hashing"; In proceeding of International Conference on Computer Vision (ICCV), 2013.

5. Xi Li*, Guosheng Lin*, Chunhua Shen, Anton van den Hengel and Anthony Dick **(* indicates equal contribution)**;
"Learning Hash Functions Using Column Generation"; In proceeding of International Conference on Machine Learning (ICML), 2013.

6. Guosheng Lin, Chunhua Shen and Anton van den Hengel;
"Approximate Constraint Generation for Efficient Structured Boosting"; In proceeding of International Conference on Image Processing (ICIP), 2013.

7. Guosheng Lin, Chunhua Shen, Anton van den Hengel and David Suter;
"Fast Training of Effective Multi-class Boosting Using Coordinate Descent Optimization"; In proceeding of Asian Conference on Computer Vision (ACCV), 2012.

THE UNIVERSITY OF ADELAIDE

Abstract

Faculty of Engineering, Computer and Mathematical Sciences
School of Computer Science

Doctor of Philosophy

by Guosheng Lin

Machine learning techniques play essential roles in many computer vision applications. This thesis is dedicated to two types of machine learning techniques which are important to computer vision: structured learning and binary code learning. Structured learning is for predicting complex structured output of which the components are inter-dependent. Structured outputs are common in real-world applications. The image segmentation mask is an example of structured output. Binary code learning is to learn hash functions that map data points into binary codes. The binary code representation is popular for large-scale similarity search, indexing and storage. This thesis has made practical and theoretical contributions to these two types of learning techniques.

The first part of this thesis focuses on boosting based structured output prediction. Boosting is a type of methods for learning a single accurate predictor by linearly combining a set of less accurate weak learners. As a special case of structured learning, we first propose an efficient boosting method for multi-class classification, which can be applied to image classification. Different from many existing multi-class boosting methods, we train class specified weak learners by separately learning weak learners for each class. We also develop a fast coordinate descent method for solving the optimization problem, in which we have closed-form solution for each coordinate update.

For general structured output prediction, we propose a new boosting based method, which we refer to as StructBoost. StructBoost supports nonlinear structured learning by combining a set of weak structured learners. Our StructBoost generalizes standard boosting approaches such as AdaBoost, or LPBoost to structured learning. The resulting optimization problem is challenging in the sense that it may involve exponentially many variables and constraints. We develop cutting plane and column generation based algorithms to efficiently solve the optimization. We show the versatility and usefulness of StructBoost on a range of problems such as optimizing the tree loss for hierarchical multi-class classification, optimizing the Pascal overlap criterion for robust visual tracking and learning conditional random field parameters for image segmentation.

The last part of this thesis focuses on hashing methods for binary code learning. We develop three novel hashing methods which focus on different aspects of binary code learning. We first present a column generation based hash function learning method for preserving triplet based relative pairwise similarity. Given a set of triplets that encode the pairwise similarity comparison information, our method learns hash functions within the large-margin learning framework. At each iteration of the column generation procedure, the best hash function is selected. We show that our method with triplet based formulation and large-margin learning is able to learn high quality hash functions.

The second hashing learning method in this thesis is a flexible and general method with a two-step learning scheme. Most existing approaches to hashing apply a single form of hash function, and an optimization process which is typically deeply coupled to this specific form. This tight coupling restricts the flexibility of the method to respond to the data, and can result in complex optimization problems that are difficult to solve. In this chapter we propose a flexible yet simple framework that is able to accommodate different types of loss functions and hash functions. This framework allows a number of existing approaches to hashing to be placed in context, and simplifies the development of new problem-specific hashing methods. Our framework decomposes the hashing learning problem into two steps: hash bit learning and hash function learning based on the learned bits. The first step can typically be formulated as binary quadratic problems, and the second step can be accomplished by training standard binary classifiers. These two steps can be easily solved by leveraging sophisticated algorithms in the literature.

The third hashing learning method aims for efficient and effective hash function learning on large-scale and high-dimensional data, which is an extension of our general two-step hashing method. Non-linear hash functions have demonstrated their advantage over linear ones due to their powerful generalization capability. In the literature, kernel functions are typically used to achieve non-linearity in hashing, which achieve encouraging retrieval performance at the price of slow evaluation and training time. We propose to use boosted decision trees for achieving non-linearity in hashing, which are fast to train and evaluate, hence more suitable for hashing with high dimensional data. In our approach, we first propose sub-modular formulations for the hashing binary code inference problem and an efficient GraphCut based block search method for solving large-scale inference. Then we learn hash functions by training boosted decision trees to fit the binary codes. We show that our method significantly outperforms most existing methods both in retrieval precision and training time, especially for high-dimensional data.

Dedicated to my family.

Contents

List of Figures

List of Tables