

Sequential Labeling with Structural SVM Under Non-decomposable Loss Functions



Guopeng Zhang

Supervisor: Prof. Massimo Piccardi

Faculty of Engineering and Information Technology
University of Technology, Sydney

This dissertation is submitted for the degree of
Doctor of Philosophy

March 2017

Declaration

I, Guopeng Zhang, hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 65,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Guopeng Zhang
March 2017

Acknowledgements

It is not easy to just say "I would like to thank someone" in the acknowledgements section of a PhD thesis. And I really lack words to express my heartfelt thanks to those who have helped me during my PhD study. However, I will always remember that I had the privilege to work with such excellent researchers.

- I would like to sincerely thank Prof. Massimo Piccardi for his kind help during the whole period of my PhD study. Prof. Massimo Piccardi constantly held meetings with me every week to give essential advice on theory and paper annotation in order to help me become a researcher. He is not only a professor but a good friend to me. Since I first came to Sydney he has cared about my life in every respect. It has been an honor for me to be his student and to work with him.
- Dr. Ava Bargi and Dr. Ehsan Zare Borzeshi are also very good researchers. They both started PhD study earlier than me and gave me helpful advice. We frequently exchanged ideas and held discussions with each other to improve our paper writing. I am very glad to have had the chance to work with them.
- I would also like to thank Dr. Shaukat Abidi, Mr. Zhen Wang, Mrs. Fairouz Hussein and Mr. Sari Awwad. Although PhD study is hard and tiring, I was always happy to chat and drink coffee with you all!
- I would like to dedicate this thesis to my loving parents Zhenbo Zhang and Meilin Zhou who supported me emotionally and financially for the duration of my time in Sydney. They made every effort to ensure that I studied and lived well here.
- Last but not least, I would like to extend my sincere thanks to my lovely wife Zhibin Li, who has been my constant support for over 11 years. She has given me more than I can ever say. She worked hard to earn money during my study and cooked meals for me. She encouraged me whenever I felt down or fatigued. Without her, I definitely would not have got through the years of exacting study and research. I thank her most

sincerely for everything that she has done for me.

Guopeng Zhang,
June 2016, Sydney

Abstract

This thesis mainly focuses on the sequential labeling problem. Sequential labeling is a fundamental problem in computer vision and machine learning areas and has been researched in many applications. The most popular model for sequential labeling is the hidden Markov model where the sequence of class labels to be predicted is encoded as a Markov chain. In recent years, other structural models, in particular, the extension of SVM to the classification of sequences and other structures have benefited from minimum-loss training approaches which in many cases lead to greater classification accuracy. However, SVM training requires the choice of a suitable loss function. Common loss functions available for training are restricted to decomposable cases such as the zero-one loss and the Hamming loss. Other useful losses such as the F_1 loss, average precision (AP) loss, equal error rates and others are not available for sequential labeling. For the average precision, some results have been proposed in the past, but our results are more general. On the other hand, classification accuracy often suffers from the uncertainty of ground truth labeling and traditional structural SVM only ensures that the ground-truth labeling of each sample receives a score higher than that of any other labeling. However, no specific score ranking is imposed among the other labelings.

For the loss functions problem, we propose a training algorithm that can cater for the F_1 loss and any other loss function based on the contingency table. In our thesis, we propose exact solutions for the F_1 loss, precision/recall at fixed value of recall/precision, precision for a fixed value of predicted positives ("precision at k"), precision/recall Break-Even Point and a formulation of the Average Precision (AP loss). For further experiments, we not only apply the AP loss in the training, but also in testing.

For the uncertainty in the ground-truth labeling problem, we extend the standard constraint set of structural SVM with constraints between "almost-correct" labelings and less desirable ones to obtain a partial ranking structural SVM (PR-SSVM) approach.

We choose different datasets to verify our approaches: human activity datasets including the challenging TUM Kitchen dataset and CMU-MMAC dataset, and the Ozone Level Detection dataset. The experimental results show the efficiency of our approaches on different performance measurements, such as detection rate, false alarm rate and F_1 measure,

compared to the conventional SVM, HMM and structural SVM with decomposable losses such as the 0-1 loss and Hamming loss.

Table of contents

| | |
|--|-------------|
| List of figures | xiii |
| List of tables | xv |
| 1 Introduction | 1 |
| 2 Literature Review | 5 |
| 2.1 Activity Recognition Review | 5 |
| 2.1.1 Taxonomy of Human Activities | 5 |
| 2.1.2 Activity Segmentation and Classification | 6 |
| 2.1.3 Commonly Used Datasets in Human Activity Recognition | 7 |
| 2.1.4 Image Representation | 12 |
| 2.1.5 Application-specific Representations | 21 |
| 2.1.6 Activity detection | 22 |
| 2.1.7 Summary of Image Representation | 23 |
| 2.2 Machine Learning Review | 25 |
| 2.2.1 Direct classification | 25 |
| 2.2.2 Dimensionality reduction | 25 |
| 2.2.3 Nearest neighbor classification | 26 |
| 2.2.4 Discriminative classifiers | 27 |
| 2.2.5 Temporal state-space models | 27 |
| 2.2.6 Dynamic time warping | 27 |
| 2.3 Joint Activity Segmentation and Classification | 28 |
| 2.4 Multi-label Classification | 31 |
| 2.4.1 Problem Transformation | 31 |
| 2.5 Sequential Labeling | 34 |
| 2.5.1 Discrete Markov Models | 35 |
| 2.5.2 Hidden Markov Models | 36 |

| | | |
|----------|--|-----------|
| 2.6 | Support Vector Machine | 43 |
| 2.6.1 | Binary SVM | 43 |
| 2.6.2 | Multi-class SVM | 53 |
| 2.6.3 | Structural SVM | 57 |
| 2.6.4 | Multi-label Classification with Arbitrary Loss Function | 65 |
| 2.6.5 | Loss Functions | 66 |
| 3 | Sequential Labeling With Structural SVM Under the Average Precision Loss | 71 |
| 3.1 | Introduction and Related Work | 71 |
| 3.2 | Background | 73 |
| 3.2.1 | Average Precision | 73 |
| 3.3 | Training and Testing Sequential Labeling with the AP Loss | 74 |
| 3.3.1 | Inference and loss-augmented inference | 75 |
| 3.4 | Experiments | 76 |
| 3.4.1 | Results on the TUM Kitchen dataset | 78 |
| 3.4.2 | Results on the CMU Multimodal Activity dataset | 79 |
| 3.5 | Conclusion | 80 |
| 4 | Sequential Labeling With Structural SVM Under Non-decomposable Losses | 81 |
| 4.1 | Augmented inference for sequential labeling under loss functions of the classification contingency table | 81 |
| 4.2 | Experiments | 85 |
| 4.2.1 | TUM Kitchen dataset | 86 |
| 4.2.2 | CMU Multimodal Activity dataset | 86 |
| 4.2.3 | Ozone Level Detection dataset | 90 |
| 4.3 | Conclusion | 90 |
| 5 | Structural SVM With Partial Ranking for Activity Segmentation and Classification | 91 |
| 5.1 | Introduction and related work | 91 |
| 5.2 | Loss Function and Partial Ranking Training | 93 |
| 5.2.1 | Loss function | 93 |
| 5.2.2 | Training by partial ranking | 93 |
| 5.3 | Extended Primal Problem | 94 |
| 5.4 | Experimental Results and Discussion | 95 |
| 5.4.1 | Results on the TUM Kitchen dataset | 96 |
| 5.4.2 | Results on the CMU-MMAC dataset | 97 |

Table of contents xi

5.5 Conclusion 98

6 Conclusion 101

References 103

Appendix A Algorithm Correctness 117

Index 119

List of figures

| | | |
|------|--|----|
| 2.1 | Different actions contained by KTH human motion dataset. | 8 |
| 2.2 | Human action dataset recorded at the Weizmann institute. | 9 |
| 2.3 | UCF sports action dataset. | 9 |
| 2.4 | Hollywood human action dataset. | 10 |
| 2.5 | INRIA XMAS multi-view dataset. | 10 |
| 2.6 | TUM kitchen dataset. | 11 |
| 2.7 | CMU MMAC dataset. | 12 |
| 2.8 | Action sequence with features. | 28 |
| 2.9 | Time sequence which has been segmented and assigned with labels a_1, \dots, a_T | 29 |
| 2.10 | Baseline approach of segmentation and classification jointly. | 30 |
| 2.11 | Our approach for segmentation and classification jointly. | 30 |
| 2.12 | Example of a multi-label dataset. | 31 |
| 2.13 | Transformation of the dataset using (a) <i>copy</i> , (b) <i>copy-weight</i> , (c) <i>select-max</i> , (d) <i>select-min</i> , (e) <i>select-random</i> (one of the possible) and (f) <i>ignore</i> | 32 |
| 2.14 | Transformed data using label powerset method. | 32 |
| 2.15 | Dataset produced by the BR method. | 33 |
| 2.16 | Dataset produced by the RPC method. | 34 |
| 2.17 | Hidden Markov Model | 37 |
| 2.18 | Binary classification problem | 45 |
| 2.19 | Classification of an unknown pattern by an SVM. The pattern is in input space compared to support vectors. The resulting values are non-linearly transformed. A linear function of these transformed values determines the output of the classifier | 46 |
| 2.20 | Geometric margin and hyperplane | 51 |
| 2.21 | Margin with outlier | 52 |

| | | |
|------|--|----|
| 2.22 | (a) The decision DAG for finding the best class out of 4 classes. The equivalent list state for each node is shown next to that node. (b) A diagram of the input space of a 4-class problem. A one-versus-one SVM can only exclude one class from consideration. | 56 |
| 2.23 | Illustration of natural language parsing model | 58 |
| 4.1 | An example of results from TUM Kitchen sequence 16, left hand, for action <i>OpeningADrawer</i> : a) structural SVM with the 0-1 loss and Hamming losses, HMM and standard SVM versus the ground truth; b) structural SVM with the F_1 loss versus the ground truth. This figure is better viewed in colour. . . | 89 |
| 5.1 | Example of detection (video 21, right hand, action <i>ClosingADoor</i>): a) Baseline; b) Structural SVM (SSVM); c) Partial-ranking structural SVM (PR-SSVM). | 97 |

List of tables

| | | |
|-----|---|----|
| 3.1 | Comparison of the average precision over the TUM Kitchen dataset. SVM: standard SVM baseline; 0-1 loss and Hamming loss: structural SVM with conventional loss functions; AP loss: proposed technique. | 78 |
| 3.2 | Comparison of the average precision over the CMU-MMAC dataset. SVM: standard SVM baseline; 0-1 loss and Hamming loss: structural SVM with conventional loss functions; AP loss: proposed technique. | 79 |
| 4.1 | Comparison of F_1 measure, DR and FAR over the TUM Kitchen dataset (left hand sequences). F_1 loss, AP loss: proposed techniques; 0-1 loss and Hamming loss: structural SVM with conventional loss functions; HMM: hidden Markov model baseline; SVM: frame-by-frame SVM baseline . . . | 87 |
| 4.2 | Comparison of F_1 measure, DR and FAR over the TUM Kitchen dataset (right hand sequences). | 87 |
| 4.3 | Comparison of F_1 measure, DR and FAR over the CMU-MMAC dataset. . . | 88 |
| 4.4 | Comparison of F_1 measure, DR and FAR over the Ozone Level Detection dataset. | 88 |
| 5.1 | Comparison of detection rate, false alarm rate and $F1$ score on the TUM Kitchen dataset (right hand). | 96 |
| 5.2 | Comparison of detection rate, false alarm rate and $F1$ score on the TUM Kitchen dataset (left hand). | 97 |
| 5.3 | Comparison of frame-level accuracy with previous results for the TUM Kitchen dataset. | 97 |
| 5.4 | Comparison of detection rate, false alarm rate and $F1$ score on the CMU-MMAC dataset (“brownies”). | 98 |
| 5.5 | Comparison of frame-level accuracy with previous results for the CMU-MMAC dataset. | 98 |

