

Continuous States Conditional Random Fields Training Using Adaptive Integration



Submitted by

João José Amaro Amador Leitão

to the

University of Exeter

as a thesis for the degree of

Doctor of Philosophy in Computer Science

December 2010

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

.....

Abstract

The extension of Conditional Random Fields (CRF) from discrete states to continuous states will help remove the limitation of the number of states and allow new applications for CRF.

In this work, our attempts to obtain a correct procedure to train continuous state conditional random fields through maximum likelihood are presented. By deducing the equations governing the extension of the CRF to continuous states it was possible to merge with the Particle Filter (PF) concept to obtain a formulation governing the training of continuous states CRFs by using particle filters. The results obtained indicated that this process is unsuitable because of the low convergence of the PF integration rate in the needed integrations replacing the summation in CRFs. So a change in concept to an adaptive integration scheme was made.

Based on an extension of the Binary Space Partition (BSP) algorithm an adaptive integration process was devised with the aim of producing a more precise integration while retaining a less costly function evaluation than PF. This allowed us to train continuous states conditional random fields with some success.

To verify the possibility of increasing the dimension of the states as a vector of continuous states a scalable version was also used to briefly assess its fitness in two-dimensions with quadtrees. This is an asymmetric two-dimensional space partition scheme.

In order to increase the knowledge of the problem it would be interesting to have further information of the relevant features. A feature selection embedded method was used based on the lasso regulariser with the intention of pinpointing the most relevant feature functions indicating the relevant features.

Contents

Acknowledgements	4
1 Introduction	10
1.1 Limitations of existing models	10
1.2 Purposed extensions	13
1.3 Novel Contributions	14
1.4 Work Structure	14
2 Background	16
2.1 Introduction	16
2.2 Sequence Classifiers	18
2.2.1 Hidden Markov Models	18
2.2.2 Conditional Random Fields	25
2.2.3 Training methods	29
2.2.4 Inference and the Viterbi algorithm	31
2.3 Continuous States and Observations	33
2.3.1 Linear Quadratic Estimator	33
2.4 Particle Filters	36
2.5 Conclusions	39
3 Maximum Likelihood of continuous state Conditional Random Fields using Monte Carlo particles	41
3.1 Introduction	41
3.2 Maximum Entropy	43
3.3 Maximum Likelihood	44

3.4	Maximum Likelihood with Monte Carlo Particles	48
3.5	Gradient of the Maximum Likelihood with Monte Carlo Particles	51
3.6	Results	55
3.7	Conclusions	62
4	Using Adaptive Integration for Learning Conditional Random Fields	64
4.1	Introduction	64
4.2	Binary Space Partition and Integration	65
4.3	CRF with Ternary Space Partition Integration	73
4.4	Regularisation	78
4.5	Results	79
4.6	Conclusions	85
5	Bi-dimensional Continuous States Conditional Random Fields	87
5.1	Introduction	87
5.2	Quadtrees	88
5.3	Adaptive Integration with Quadtrees	89
5.4	Dimension extension on the Adaptive Integration Concept in CRF	91
5.5	Results	92
5.6	Conclusion	96
6	Feature Functions Selection for Continuous State Conditional Random Fields	98
6.1	Introduction	98
6.2	Parameter shrinkage	100
6.3	Lasso Regulariser in Continuous State Conditional Random Fields	102
6.4	Results	103
6.5	Conclusion	106
7	Summary, Conclusions and Future Work	108
7.1	Summary	108
7.2	Conclusions	110
7.3	Future Work	111

List of Figures

2.1	Hidden Markov Model transitions relation relations	19
2.2	Conditional Random Fields transition relations diagram	28
3.1	#1 set of feature functions with $\lambda_0 = 1$	57
3.2	#1 set of feature functions with $\lambda_0 = 10$	57
3.3	#2 set of feature functions with $\lambda_0 = 10$	58
3.4	Alpha and Beta pass re-normalised potentials from Riemann and Particle Filter integration for $t = 10, 11$	59
3.5	Comparison between PF and Riemann integration for a range of λ_1	61
4.1	Example of a TSP partition.	69
4.2	Example of a local transition potential.	80
4.3	Observations and results on the state space.	81
4.4	Trajectory of the parameters in the parameters space.	82
4.5	Negative log-likelihood and gradient for TSP and Riemann integration over a range of lengths of the training sequence.	83
4.6	Additive inverse of the error function and its gradient.	84
5.1	Example of a asymmetric partition of a square using quadtrees.	88
5.2	Observations and results in the bi-dimensional state space.	94
5.3	Projections of the observations and results in to the state-time planes and state-state plane.	95
6.1	Examples of $\ x\ _p = 1$ for several values of p.	101
6.2	Value of the function $-\log \lambda$ for each feature function for problem #1 . . .	106
6.3	Value of the function $-\log \lambda$ for each feature function for problem #2 . . .	107

List of Tables

3.1	Set #1 of feature functions for testing.	56
3.2	Set # 2 of feature functions for testing.	56
4.1	Results for the TSP and Riemann integration test.	70
5.1	Feature functions for the bi-dimension testing problem.	93
6.1	Problems for the parameters shrinkage testing.	104
6.2	Feature functions for the parameters shrinkage testing.	105
6.3	Parameters for the feature functions time shift.	105