

Contents

1. Introduction	1
1.1. Scope	2
1.2. Research Questions and Thesis Structure	3
1.3. List of Publications	5
1.3.1. Abstractions Research	5
1.3.2. Auxiliary Publications	9
2. Foundations	11
2.1. Models for Sequential Decision Making	11
2.1.1. Markov Decision Processes	11
2.1.2. Multi-agent MDPs	14
2.1.3. Partially Observable Stochastic Games	15
2.2. Monte Carlo Tree Search	16
2.2.1. Directed Acyclic Graph MCTS	19
2.2.2. Multi-agent MCTS	19
2.3. Abstractions	20
2.3.1. Benefits of Abstractions for MCTS-based Methods	21
2.3.2. State-Action Pair Abstractions	22
2.3.3. State Abstractions	24
2.4. On-the-Go-Abstractions	29
2.4.1. Motivation	29
2.4.2. OGA-UCT	29
2.4.3. Parameters of OGA-UCT	29
2.4.4. Source of Abstraction Errors in OGA-UCT	31
2.4.5. ASAP-UCT and AS-UCT	31
2.4.6. RANDOM-OGA	31
3. Related Work	33
3.1. Overview	33
3.1.1. Scope	34
3.2. Abstraction Construction Methods	35
3.2.1. Online Abstraction Construction	35
3.2.2. Lossless Abstractions	37
3.2.3. Decision-Point-Dependent Abstractions	37
3.2.4. Property-based Aggregation	38
3.2.5. Model-based Abstractions	39

3.2.6.	Temporal Action Abstractions	40
3.2.7.	Multi-abstractions	40
3.2.8.	Abstractions for Stochastic Environments	41
3.2.9.	Abstractions for Combinatorial Action Spaces	42
3.2.10.	Abstractions for Continuous Action Spaces	43
3.2.11.	Abstractions for Non-cooperative Multi-agent Problems	44
3.2.12.	Abstractions for Partially Observable Domains	45
3.3.	Applications of Abstractions	46
3.3.1.	Abstraction Mapping Graphs	46
3.3.2.	Fully Abstracted Optimization and Search	47
3.3.3.	MCTS Tree Policy Improvement	48
3.3.4.	Improved alpha-beta Search Pruning	49
3.3.5.	State Evaluation Improvement	49
3.4.	Conclusion	50
4.	Default Experiment Setup	51
5.	Dynamic Exploration Factors	55
5.1.	λ -strategies From the Literature	56
5.2.	Formalization of New and Old λ -strategies	57
5.3.	Experiments	61
5.3.1.	Parameter-optimized Performances	61
5.3.2.	Pairings and Relative Improvement Scores	64
5.3.3.	Ablation: Performances on Same Reward Scale Environments	66
5.4.	Summary and Conclusion	68
5.5.	Limitations and Future Work	68
6.	Abstraction Building	71
6.1.	Problems of the ASAP Framework	71
6.2.	An Existing OGA-UCT Modification: Pruned OGA	74
6.3.	Integrating Approximate Abstractions with OGA: (ϵ_a, ϵ_t) -OGA	76
6.3.1.	(ϵ_a, ϵ_t) -OGA	76
6.3.2.	Experiments	81
6.3.3.	Summary and Conclusion	90
6.3.4.	Limitations and Future Work	91
6.4.	Known Value Difference Abstractions	92
6.4.1.	Which Abstractions ASAP Misses	92
6.4.2.	The Known-Value-Difference-Abstractions Framework	94
6.4.3.	ϵ_t -KVDA, and α -KVDA	95
6.4.4.	Experiments	98
6.4.5.	Summary and Conclusion	105
6.4.6.	Limitations and Future Work	105
6.5.	State Abstractions Through Action Pruning	106
6.5.1.	Why ASAP Misses Some State Equivalences	106

6.5.2.	Ideal Pruning Abstractions	107
6.5.3.	IPA-UCT	108
6.5.4.	Experiments	111
6.5.5.	Summary and Conclusion	119
6.5.6.	Limitations and Future Work	119
6.6.	Statistical Abstractions	121
6.6.1.	How Action Equivalences Can Be Detected in SysAdmin	121
6.6.2.	Layerwise Coinciding Reward Distributions Abstraction	123
6.6.3.	Approximating LCRDA Abstractions: AUPO	124
6.6.4.	Experiments	131
6.6.5.	Summary and Conclusion	138
6.6.6.	Limitations and Future Work	139
7.	Abstraction Usage	141
7.1.	Abstraction Dropping	141
7.1.1.	Reasons for Abstraction Dropping	141
7.1.2.	Opportunities for Improvement in Existing Abstraction Dropping Methods	142
7.1.3.	Confidence-based Abstraction Dropping	144
7.1.4.	Experiments	145
7.1.5.	Summary and Conclusion	154
7.1.6.	Limitations and Future Work	154
7.2.	Intra-abstraction Policies	155
7.2.1.	Introducing Different Intra-abstraction Policies	155
7.2.2.	Case Study: RANDOM vs UCT Intra-abstraction Policy	157
7.2.3.	Experiments	158
7.2.4.	Summary and Conclusion	167
7.2.5.	Limitations and Future Work	167
8.	Summary and Conclusion	169
9.	Future Research Directions	171
A.	Appendix	173