1 Introduction

My main research interests lie in the use and development of *planning, learning and sensing* technologies for autonomous agents and multi-agent systems, particularly in the field of robotics, such as robot soccer and intelligent service robots. My research falls mostly into one or more of the following topics: decision-theoretical planning, multi-agent systems, reinforcement learning, multi-human tracking and human-robot interaction.

In the research of Artificial Intelligence, agent-based paradigm aims to provide a unifying framework for conceptualizing, designing, and implementing intelligent systems, that sense, act and learn autonomously in dynamic and/or stochastic environments, to solve a growing number of complex problems. Agents, particularly various kinds of robots, are playing more and more important roles in world economics and people's everyday life, from satellites to smartphones. Generally speaking, perceptional inputs from sensors have inevitable noises and errors. The effects of actuators have also unpredictable impact with noises, or even failures. There may also exist different levels of hidden information that can not be observed directly.

My research aims to enable agents the abilities of sensing, planing and learning under uncertainties in principled ways. My work mainly follows a decision-theoretic framework called partially observable Markov decision processes (POMDPs) [22], which model sensing, planning and learning in a Bayesian-optimal way. Specifically, we propose a novel algorithm that addresses the key bottleneck of online planning in large MDPs using MAXQ [16] based hierarchical decomposition [1, 6, 5, 8]; we apply the idea of Thompson sampling [32] in Monte-Carlo planning and learning for MDPs and POMDPs [7, 9]; we implement an intelligent service system, with multiple cooperative robots, that serves people in a bar environment [25]; and, we develop a particle filtering over sets (PFS) approach to intention-aware multi-human tracking in the domain of human-robot interaction [3].

In future research, I mainly look forward to being in part of the development and application of principled methods that can tackle increasing complexities of future systems of Artificial Intelligence in terms of planning, learning and sensing in dynamic and/or stochastic environments.

2 Current Projects

2.1 Meta-reasoning Approach for Real-Time Online Planning, 2015

In dynamic environment, the environment may change before the agent has actually acted. To successfully interact with such environment, the agent has to act reactively with limited deliberation time of each decision cycle, otherwise the environment will certainly be out of control, may even leading to some catastrophic consequences. On the other hand, meaningful results usually can only be obtained by enough deliberation. Thus the agent has to conduct continuous computation over multiple decision cycles to deliberate sufficiently, for example, by searching over game trees or iterating utility functions. The tradeoff between reactive and deliberative computations is then the key to success for real-time online planning in dynamic environment. A meta-reasoning approach is developed to tackle this funder mental problem. In the object level, the agent simultaneously computes for two policies: one global policy which intends to solve the original optimization problem; one local policy which intends to ensure that the agent can have more time to compute by acting safe when interacting with the environment. The difference between local and global polices lies in the fact that local policy is using a different reward function from the global policy. In the meta-level, a reactive process is used to quickly evaluate which policy to follow at the beginning of each decision cycle. The evaluation is based on meta-level utility functions for these two policies. In experiments, the meta-reasoning approach can successfully complete a real-time race-track problem (with collision) with 100% probability, while conventional algorithms fail to do so.

3 Past Projects

3.1 Semi-Markov Process based User Behavior Modelling, 2015

A classification system is built based on the theory of semi-Markov process to detect whether a potential buyer is involved in cheating when she/he is buying items online via Taobao.com and Tmall.com. A buyer is cheating if she/he is tasked by the seller to buy specified items to actually increase the selling amount and/or credit history of that seller. This type of cooperative cheating behaviour between sellers and buyers is rather harmful to maintain a reliable and trustful ecosystem for online shopping service, which is also very difficult to be detected. A semi-Markov based classification system is developed to estimate the cheating probability of particular users. A full online shopping behaviour is separated into different stages, including searching, browsing, paying, shipping and reviewing. Within each stage, user behaviours in terms of clicking on webpages are recognized as different states. State and stage transitions are then learned from recorded user behaviour history labelled with different cheating classes. The final transition matrix also includes distributions on time durations between linked events. The overall semi-Markov process contains more than 500 nodes. One assumption for this work is that normal and cheating shopping behaviours have clearly different transition matrices. In our testing data, the final system can correctly detect a user's cheating behaviour with precision of 0.91 and recall of 0.86.

3.2 Multi-Human Tracking and Intention-Recognition, 2013 - 2014

The ability for an autonomous robot to track and identify multiple humans and understand their intentions is crucial for socialized human-robot interactions in dynamic environments. Take CoBot [27] trying to enter an elevator as an example. When the elevator door opens, suppose there are multiple humans occupied, CoBot needs to track each human's state and intention in terms of whether he/she is going to exit the elevator or not. For the purposes of safely and friendly interacting with humans, CoBot can only make the decision to enter the elevator when any human who intends to exit is believed to have exited.

We propose a novel particle filtering over sets (PFS) approach, together with associated techniques we introduce to make PFS possible, including: 1) the assignment and false-missing pruning strategies to approximate the observation function, 2) a data-association based particle refinement method, 3) a Bayesian density estimation approach to estimate motion and proposal weights, and 4) an expectation-maximization (EM) based human identification process to recognize each individual human from particles. Intentions are recognized by associating different motion models for different intentions. In multi-object tracking (MOT) domain, most existing

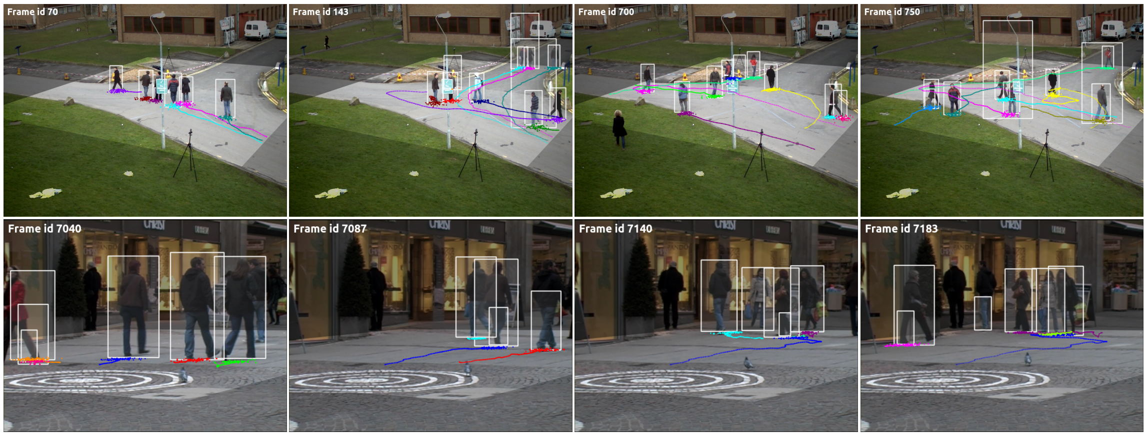

Figure 1: (Best viewed in color and zoom-in.) Tracking results of PFS in PETS2009 S2L1 and TUD-Stadtmitte dataset. White bounding boxes are the raw detections. States of particles and trajectories are depicted with different colors indicating different identified humans. A short video showing the whole results compared with ground truth in PETS2009 dataset is available at `http://goo.gl/35UOy1`.

approaches assume one or more hypotheses of data associations between observations and targets, and apply Bayesian filtering on each target separately [19, 26, 33, 12, 19]. It is difficult for these methods to recover from wrong assumptions. Our approach avoids directly performing observation-to-target association, by using a joint state to encode the entire multi-target state including the number of targets, the state of each target, and implicitly all possible hypotheses. The overall method outperforms the state-of-the-art in the challenging PETS2009 dataset [18] in terms of CLEAR MOT metrics [11]. The real robot experiments indicate that a robot integrated with this approach is able to track random humans in real time, and understand their intentions in terms of *moving* and *staying*. Figure 1 shows some tracking examples. A video showing CoBot following a human for approximately 10 minutes based on the results provided by PFS can also be found at `http://goo.gl/T3UtlS`. Related publications are listed as follows.

1. **Aijun Bai**, Reid Simmons, Manuela Veloso, and Xiaoping Chen. Intention-aware multi-human tracking for human-robot interaction via particle filtering over sets. In *AAAI 2014 Fall Symposium: AI for Human-Robot Interaction (AI-HRI 2014)*, Arlington, United States, 2014 (accepted)

3.3 Bayesian Monte-Carlo Tree Search with Thompson Sampling, 2012 - 2013

Monte-Carlo tree search (MCTS) has been drawing great interest recently in domains of planning and learning under uncertainty [23, 24, 21, 14, 17]. One of the key challenges is the trade-off between exploration and exploitation. To address this, we introduce novel approaches to MCTS by using posterior sampling to select actions for Monte-Carlo online planning in the contexts of MDPs and POMDPs. We propose the DNG-MCTS and D^{2}NG-POMCP algorithms – novel Bayesian approaches for online planning in MDPs and POMDPs within MCTS framework by applying Thompson sampling as the action selection strategy. Specifically, we apply MCTS to MDPs and POMDPs, and treat the cumulative reward returned by taking an action from a search node in the MCTS search tree as a random variable following an unknown distribution. We parametrize the distribution by introducing necessary hidden parameters, and infer the posterior distribution of the hidden parameters in Bayesian settings by appropriately choosing a conjugate prior. Thompson sampling is then used to exploit and explore the search tree, by selecting an action based on its posterior probability of being optimal. We show that the proposed algorithms are guaranteed to converge to the near-optimal policy in the search tree at infinity. Experimental results in MDPs confirm that, comparing to the general UCT algorithm [24], DNG-MCTS produces competitive results in the CTP domain, and converges faster in the domains of racetrack and sailing with respect to sample complexity [13]. Experimental results in POMDPs show that D^2NG-POMCP outperforms the state-of-the-art algorithms in both RockSample and PocMan domains [30]. Related work is published in **NIPS** 2013 and **ICAPS** 2014.

1. **Aijun Bai**, Feng Wu, Zongzhang Zhang, and Xiaoping Chen. Thompson sampling based Monte-Carlo planning in POMDPs. In *Proceedings of the 24th International Conference on Automated Planning and Scheduling (ICAPS 2014)*, Portsmouth, United States, 2014
2. **Aijun Bai**, Feng Wu, and Xiaoping Chen. Bayesian mixture modelling and inference based Thompson sampling in Monte-Carlo tree search. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 1646–1654. 2013

3.4 An Intelligent Service System with Multiple Robots, 2013

A multi-robot project where an intelligent bartender-waiter robot system serves people in a bar. In this system, KeJia robot [15] plays the role of bartender who recognizes and grasps the drink by following the order of people and gives it to a TurtleBot, several TurtleBots deliver drinks to people as waiters. My main work focuses on autonomous navigation and multi-robots collision avoidance for TurtleBots. We use Kinect to simulate the laser input, GMapping algorithm for SLAM, and Dijkstra algorithm and dynamic window approach (DWA) for global and local path planning respectively. For the multi-robots collision avoidance problem, our basic idea is that each robot broadcasts its own positions to other robots, and maintains other robots' position as dynamic obstacles in its own local map. Dijkstra algorithm based re-planning method is used to re-generate path plan to avoid collisions if any future collision is detected. We have participated in the Robot Competition of IJCAI 2013.

1. Qiang Lu, Guanghui Lu, **Aijun Bai**, Dongxiang Zhang, and Xiaoping Chen. An intelligent service system with multiple robots. In *Robot Competition of International Joint Conference on Artificial Intelligence (IJCAI 2013)*, Beijing, China, 2013

3.5 Hierarchical Online Planning for Large MDPs, 2010 - 2012

We developed a hierarchical online planning algorithm, namely MAXQ-OP, that benefits from the advantage of hierarchical decomposition. It recursively expands the search tree online and searches over the policy space by following the underlying task hierarchy. This is efficient since only relevant states and actions are considered according to the MAXQ hierarchy. Another contribution of this work is the completion function approximation method which make it possible

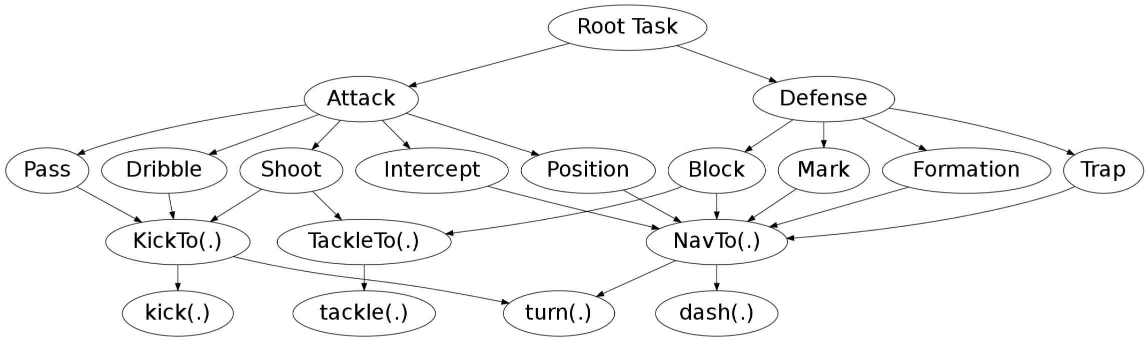

Figure 2: MAXQ task graph for WrightEagle.

to be computed online. The key observation is that the terminating distribution is relatively easy to be approximated either online or offline given domain knowledge. The empirical results show that MAXQ-OP is able to find a near-optimal policy online in the Taxi domain [16]. We have also been conducting a long-term case-study with the RoboCup soccer simulation 2D domain [31], which is extremely larger than domains usually studied in the literature, as the major benchmark to this research. The case-study showed that the agents developed with this framework and the related techniques reached outstanding performances, showing its high scalability to very large domains, while utilizing task hierarchy. The MAXQ task graph we developed in our RoboCup 2D team – WrightEagle – is shown in Figure 2. Some of the results have been published in **RoboCup Symposium** 2011, **AAMAS** 2012, ARMS 2012 and **RoboCup Symposium** 2012.

1. **Aijun Bai**, Feng Wu, and Xiaoping Chen. Towards a principled solution to simulated robot soccer. In Xiaoping Chen, Peter Stone, Luis Enrique Sucar, and Tijn Van der Zant, editors, *RoboCup-2012: Robot Soccer World Cup XVI*, volume 7500 of *Lecture Notes in Artificial Intelligence*. Springer Verlag, Berlin, 2013

2. **Aijun Bai**, Feng Wu, and Xiaoping Chen. Online planning for large MDPs with MAXQ decomposition (extended abstract). In *Proc. of 11th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, June 2012

3. **Aijun Bai**, Feng Wu, and Xiaoping Chen. Online planning for large MDPs with MAXQ decomposition. In *Proc. of the Autonomous Robots and Multirobot Systems workshop (at AAMAS 2012)*, Jun 2012

4. **Aijun Bai**, Xiaoping Chen, Patrick MacAlpine, Daniel Urieli, Samuel Barrett, and Peter Stone. Wright Eagle and UT Austin Villa: RoboCup 2011 simulation league champions. In Thomas Roefer, Norbert Michael Mayer, Jesus Savage, and Uluc Saranli, editors, *RoboCup-2011: Robot Soccer World Cup XV*, volume 7416 of *Lecture Notes in Artificial Intelligence*. Springer Verlag, Berlin, 2012

3.6 WrightEagle Soccer Simulation 2D Team, 2007 - 2013

I have been working for WrightEagle soccer simulation 2D team on multi-agent decision-making and real-time game play from 2007 to 2013, and have been the team leader since 2010. During

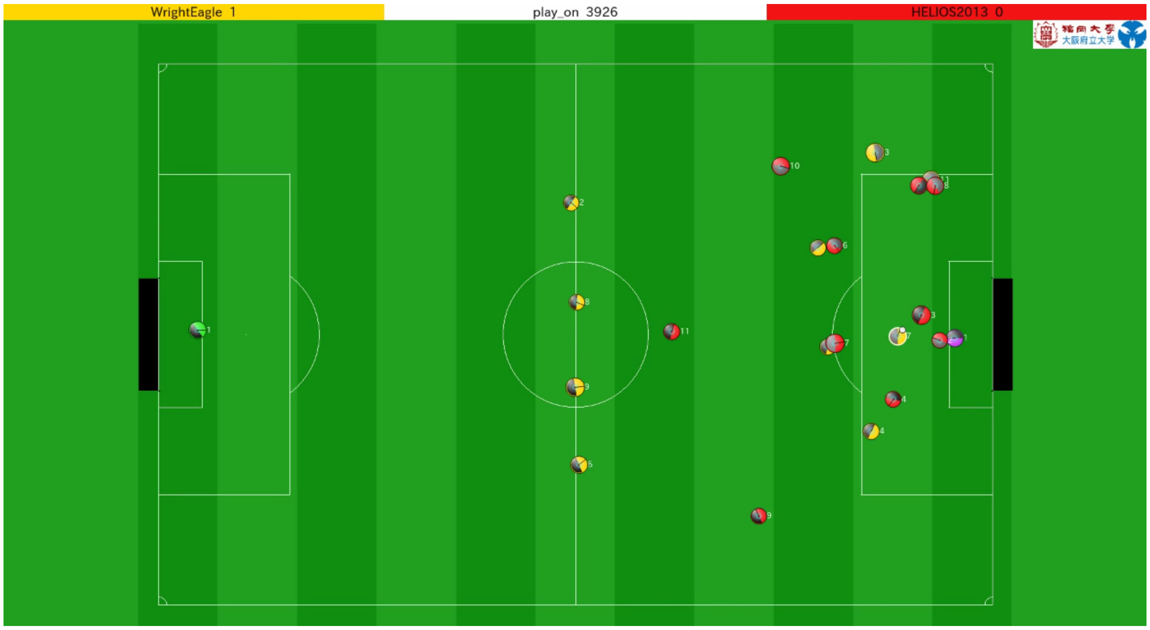

Figure 3: A snapshot of RoboCup 2013 Soccer 2D final between WrightEagle and Helios.

my working time, we have won 3 **World Champions** (2009, 2011 and 2013) and 4 runners-up (2007, 2008, 2010 and 2012) of annual RoboCup competitions, and 5 national champions (2007, 2009, 2010, 2011 and 2012) and 1 runner-up (2008) of RoboCup China Open competitions. The team website is `http://wrighteagle.org/2d/`. A video showing the final match of RoboCup 2013, between the winner – our team WrightEagle and the runner-up from Japan – Helios, can be found at `http://goo.gl/v5OMWN`. A snapshot is also shown in Figure 3.

As one of the oldest leagues in RoboCup, the soccer simulation 2D league has achieved great successes and inspired many researchers all over the world to engage themselves in this game each year [20]. Comparing to other leagues in RoboCup, the key feature of RoboCup Simulation 2D is the abstraction made by the simulator, which relieves the researchers from having to handle low-level robot problems such as object recognition, communications, and hardware issues. The abstraction enables researchers to focus on high-level functions such as planning, learning and cooperation. Detailedly, a central *server* simulates a 2-dimensional soccer field in real-time. Two teams of fully autonomous agents connect to the server via network sockets to play a soccer game over 6,000 steps (also known as *cycles*). Each team consists of 11 soccer player agents, each of which interacts independently with the server by 1) receiving a set of observations; 2) making a decision; and 3) sending actions back to the server. Observations for each player only contain noisy and local geometric information such as the distance and angle to other players, ball, and landmarks within its view range. Actions are atomic commands such as turning the body (or neck) to an angle, dashing in a given direction with certain power, kicking the ball to an angle with power, etc. The key challenge lies in the fact that it is a fully distributed, multi-agent stochastic domain with continuous state, action and observation space [31]. More details about

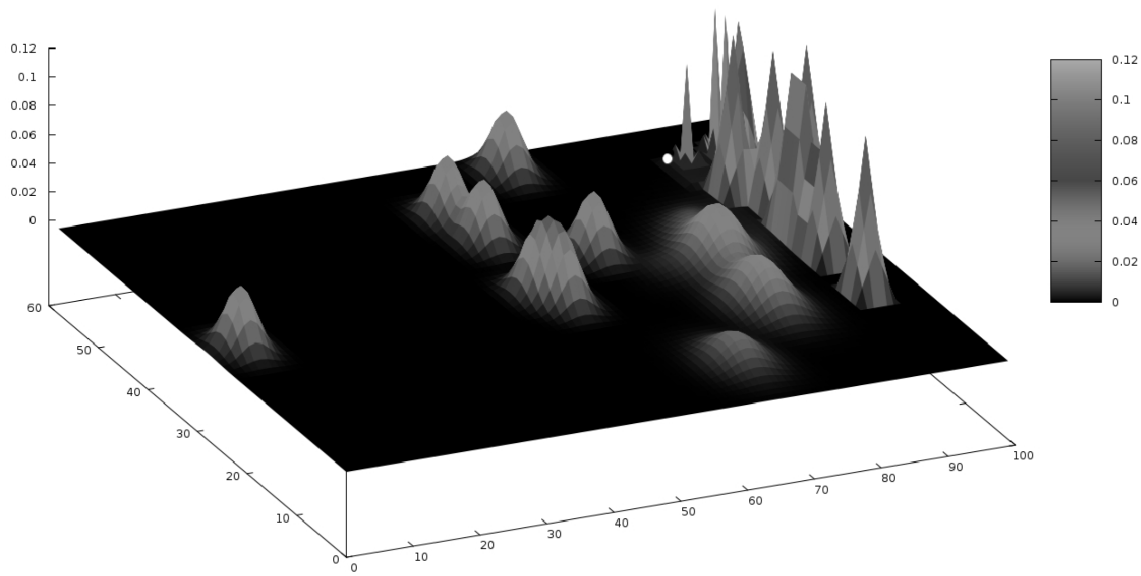

Figure 4: An example of joint belief in terms of position distributions in RoboCup 2D domain.

RoboCup Simulation 2D are available in the official website.[1]

My main work basically focuses on the entire team, particularly including Overall Decision Architecture, Behavior Attack (Probability and Evaluation System, Behavior Intercept, Behavior Shoot, Behavior Pass and Behavior Dribble), Behavior Defense (Task Allocation, Behavior Block, Behavior Mark and Behavior Formation), Behavior Goalie, Intercepting Model, Control Segment Model, A* based Motion Planning, Monte Carlo Localization and Tracking, Belief State based Sensing System, Dynamic Formation System, Communication System, etc. As an example, Figure 4 shows the joint belief state in terms of position distributions obtained from particle filters, enabling the player to track up to 10 teammates and 11 opponents with imperfect observations containing noises and errors. The core of team is mainly developed as a probabilistic planning and evaluation system within the framework of MAXQ-OP [8, 6, 5, 1] – one work of my thesis research. Besides, I have also developed the automatic test system for our team, namely AutoTest, which supports fully distributed and parallel testing of our team binaries versus other teams with web pages showing game and analysis results. Some of the base code can be publicly accessed as open source software at `https://launchpad.net/wrighteaglebase`. Historical results of WrightEagle in RoboCup annual competitions in my working period is shown in Table 1. A set of Team Description Papers for the competition is listed as follows.

1. Haochong Zhang, Miao Jiang, Haibo Dai, **Aijun Bai**, and Xiaoping Chen. WrightEagle 2D soccer simulation team description 2013. In *RoboCupg, Eindhoven, The Netherlands*, 2013

2. **Aijun Bai**, Haochong Zhang, Guanghui Lu, Miao Jiang, and Xiaoping Chen. WrightEagle 2D soccer simulation team description 2012. In *RoboCup, Mexico City, Mexico*, 2012

[1] `http://wiki.robocup.org/wiki/Soccer_Simulation_League`

Competitions	Games	Points	Goals	Win	Draw	Lost	Avg. Points	Avg. Goals
RoboCup 2007	14	34	125 : 9	11	1	2	2.42	8.92 : 0.64
RoboCup 2008	16	40	74 : 18	13	1	2	2.50	4.63 : 1.13
RoboCup 2009	14	36	81 : 17	12	0	2	2.57	5.79 : 1.21
RoboCup 2010	13	33	123 : 7	11	0	2	2.54	9.47 : 0.54
RoboCup 2011	12	36	151 : 3	12	0	0	3.00	12.6 : 0.25
RoboCup 2012	21	58	104 : 18	19	1	1	2.76	4.95 : 0.86
RoboCup 2013	19	53	104 : 9	17	2	0	2.79	5.47 : 0.47

Table 1: Historical results of WrightEagle in RoboCup annual competitions from 2007 to 2013.

3. **Aijun Bai**, Guanghui Lu, Haochong Zhang, and Xiaoping Chen. WrightEagle 2D soccer simulation team description 2011. In *RoboCup, Istanbul, Turkey*, 2011

4. **Aijun Bai**, Jing Wang, Guanghui Lu, Yuhang Wang, Haochong Zhang, Yuancong Zhu, Ke Shi, and Xiaoping Chen. WrightEagle 2D soccer simulation team description 2010. In *RoboCup, Singapore, Singapore*, 2010

5. Ke Shi, **Aijun Bai**, Yunfang Tai, and Xiaoping Chen. Wrighteagle 2009 2D soccer simulation team description paper. In *RoboCup, Graz, Austria*, 2009

6. Ke Shi, Tengfei Liu, **Aijun Bai**, Wenkui Wang, Changjie Fan, and Xiaoping Chen. WrightEagle 2008 simulation 2D team description paper. In *RoboCup, Suzhou, China*, 2008

3.7 Reversi Game, 2006

I develop a C/S framework software to play Reversi game (a.k.a. Othelo), which supports computer-computer, computer-human and human-human playing. An online-search based AI client is also well developed, with some techniques including alpha-beta pruning, history tables, transposition tables and case-based learning. The project is now publicly available at http://sourceforge.net/projects/reversigame/.

4 Future Research

For future research, I am open to all problems related to planning, sensing and learning. A few of my specific interests are described below.

I plan to model Monte-Carlo planning in a Bayesian framework, that is to determine the best sampling sequence in a Bayesian way under the constraint that one can sample from a simulator for at most N times (or in T computation time), which may be in principle suitable for addressing the exploration-exploitation challenge in MCTS. I am interested in applying Monte-Carlo planning methods in real-world problems, for example robotics task and motion planning. I am interested in utilizing POMDP framework in Bayesian reinforcement learning problems to let the agent actively learn in an optimal way, particularly in large-scale real-world applications. I am also interested in scaling up MDP and POMDP methods to very large problems by utilizing sampling and symbolic techniques.

I plan to enable robots the ability of planning and learning in belief space, for example a robot follows a team of humans by planning in belief space that can deal with hidden information, or a robot interacts with humans by having conversations and updating its internal belief gradually in order to accomplish a user task which is not clear initially due to lack of information. I am interested in handling the tradeoff between normal actions and sensing actions, that is to decide when/how to sense and when/how to act given past observation and action history. I plan to develop more robust multi-human tracking and intention-recognition algorithms in possible crowded environments, enabling robots to interact with potentially multi-humans, and understand their intentions in complex social tasks, which I believe is essential to make robots smart and safe in socialized human-robot interaction domain.

In summary, I mainly look forward to being in part of the development and application of principled methods that can tackle increasing complexities of future systems of Artificial Intelligence in terms of planning, learning and sensing in dynamic and/or stochastic environments.

References

[1] **Aijun Bai**, Xiaoping Chen, Patrick MacAlpine, Daniel Urieli, Samuel Barrett, and Peter Stone. Wright Eagle and UT Austin Villa: RoboCup 2011 simulation league champions. In Thomas Roefer, Norbert Michael Mayer, Jesus Savage, and Uluc Saranli, editors, *RoboCup-2011: Robot Soccer World Cup XV*, volume 7416 of *Lecture Notes in Artificial Intelligence*. Springer Verlag, Berlin, 2012.

[2] **Aijun Bai**, Guanghui Lu, Haochong Zhang, and Xiaoping Chen. WrightEagle 2D soccer simulation team description 2011. In *RoboCup, Istanbul, Turkey*, 2011.

[3] **Aijun Bai**, Reid Simmons, Manuela Veloso, and Xiaoping Chen. Intention-aware multi-human tracking for human-robot interaction via particle filtering over sets. In *AAAI 2014 Fall Symposium: AI for Human-Robot Interaction (AI-HRI 2014)*, Arlington, United States, 2014 (accepted).

[4] **Aijun Bai**, Jing Wang, Guanghui Lu, Yuhang Wang, Haochong Zhang, Yuancong Zhu, Ke Shi, and Xiaoping Chen. WrightEagle 2D soccer simulation team description 2010. In *RoboCup, Singapore, Singapore*, 2010.

[5] **Aijun Bai**, Feng Wu, and Xiaoping Chen. Online planning for large MDPs with MAXQ decomposition. In *Proc. of the Autonomous Robots and Multirobot Systems workshop (at AAMAS 2012)*, Jun 2012.

[6] **Aijun Bai**, Feng Wu, and Xiaoping Chen. Online planning for large MDPs with MAXQ decomposition (extended abstract). In *Proc. of 11th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, June 2012.

[7] **Aijun Bai**, Feng Wu, and Xiaoping Chen. Bayesian mixture modelling and inference based Thompson sampling in Monte-Carlo tree search. In *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 1646–1654. 2013.

[8] **Aijun Bai**, Feng Wu, and Xiaoping Chen. Towards a principled solution to simulated robot soccer. In Xiaoping Chen, Peter Stone, Luis Enrique Sucar, and Tijn Van der Zant, editors, *RoboCup-2012: Robot Soccer World Cup XVI*, volume 7500 of *Lecture Notes in Artificial Intelligence*. Springer Verlag, Berlin, 2013.

[9] **Aijun Bai**, Feng Wu, Zongzhang Zhang, and Xiaoping Chen. Thompson sampling based Monte-Carlo planning in POMDPs. In *Proceedings of the 24th International Conference on Automated Planning and Scheduling (ICAPS 2014)*, Portsmouth, United States, 2014.

[10] **Aijun Bai**, Haochong Zhang, Guanghui Lu, Miao Jiang, and Xiaoping Chen. WrightEagle 2D soccer simulation team description 2012. In *RoboCup, Mexico City, Mexico*, 2012.

[11] Keni Bernardin and Rainer Stiefelhagen. Evaluating multiple object tracking performance: the CLEAR MOT metrics. *EURASIP Journal on Image and Video Processing*, 2008, 2008.

[12] Samuel Blackrnan and Artech House. Design and analysis of modern tracking systems. *Boston, MA: Artech House*, 1999.

[13] Blai Bonet and Hector Geffner. Action selection for MDPs: Anytime AO* vs. UCT. In *AAAI Conference on Artificial Intelligence*, pages 1749–1755, 2012.

[14] Cameron Browne, Edward J. Powley, Daniel Whitehouse, Simon M. Lucas, Peter I. Cowling, Philipp Rohlfshagen, Stephen Tavener, Diego Perez, Spyridon Samothrakis, and Simon Colton. A survey of Monte Carlo tree search methods. *IEEE Trans. Comput. Intellig. and AI in Games*, 4(1):1–43, 2012.

[15] Xiaoping Chen, Jianmin Ji, Jiehui Jiang, Guoqiang Jin, Feng Wang, and Jiongkun Xie. Developing high-level cognitive functions for service robots. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, pages 989–996. International Foundation for Autonomous Agents and Multiagent Systems, 2010.

[16] Thomas G Dietterich. The maxq method for hierarchical reinforcement learning. *Proceedings of the fifteenth international conference on machine learning (1998) Volume: 8, Issue: c, Publisher: Morgan Kaufmann, Pages: 118-126*, 1999.

[17] Zohar Feldman and Carmel Domshlak. Simple regret optimization in online planning for Markov decision processes. *arXiv preprint arXiv:1206.3382*, 2012.

[18] J Ferryman and A Shahrokni. PETS2009: Dataset and challenge. In *2009 Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS-Winter)*, pages 1–6. IEEE, 2009.

[19] Thomas E Fortmann, Yaakov Bar-Shalom, and Molly Scheffe. Sonar tracking of multiple targets using joint probabilistic data association. *Oceanic Engineering, IEEE Journal of*, 8(3):173–184, 1983.

[20] T. Gabel and M. Riedmiller. On progress in robocup: the simulation league showcase. *RoboCup 2010: Robot Soccer World Cup XIV*, pages 36–47, 2011.

[21] S. Gelly and D. Silver. Monte-Carlo tree search and rapid action value estimation in computer go. *Artificial Intelligence*, 175(11):1856–1875, 2011.

[22] Leslie Pack Kaelbling, Michael L Littman, and Anthony R Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1-2):99–134, 1998.

[23] M. Kearns, Y. Mansour, and A.Y. Ng. A sparse sampling algorithm for near-optimal planning in large Markov decision processes. In *Proceedings of the 16th international joint conference on Artificial intelligence-Volume 2*, pages 1324–1331. Morgan Kaufmann Publishers Inc., 1999.

[24] L. Kocsis and C. Szepesvári. Bandit based Monte-Carlo planning. In *European Conference on Machine Learning*, pages 282–293, 2006.

[25] Qiang Lu, Guanghui Lu, **Aijun Bai**, Dongxiang Zhang, and Xiaoping Chen. An intelligent service system with multiple robots. In *Robot Competition of International Joint Conference on Artificial Intelligence (IJCAI 2013)*, Beijing, China, 2013.

[26] Donald B Reid. An algorithm for tracking multiple targets. *Automatic Control, IEEE Transactions on*, 24(6):843–854, 1979.

[27] Stephanie Rosenthal, Joydeep Biswas, and Manuela Veloso. An effective personal mobile robot agent through symbiotic human-robot interaction. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, pages 915–922. International Foundation for Autonomous Agents and Multiagent Systems, 2010.

[28] Ke Shi, **Aijun Bai**, Yunfang Tai, and Xiaoping Chen. Wrighteagle 2009 2D soccer simulation team description paper. In *RoboCup, Graz, Austria*, 2009.

[29] Ke Shi, Tengfei Liu, **Aijun Bai**, Wenkui Wang, Changjie Fan, and Xiaoping Chen. WrightEagle 2008 simulation 2D team description paper. In *RoboCup, Suzhou, China*, 2008.

[30] D. Silver and J. Veness. Monte-Carlo planning in large POMDPs. In *Advances in Neural Information Processing Systems*, pages 2164–2172, 2010.

[31] P. Stone. *Layered learning in multiagent systems: A winning approach to robotic soccer*. The MIT press, 2000.

[32] William R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25:285–294, 1933.

[33] Alper Yilmaz, Omar Javed, and Mubarak Shah. Object tracking: A survey. *Acm computing surveys (CSUR)*, 38(4):13, 2006.

[34] Haochong Zhang, Miao Jiang, Haibo Dai, **Aijun Bai**, and Xiaoping Chen. WrightEagle 2D soccer simulation team description 2013. In *RoboCupg, Eindhoven, The Netherlands*, 2013.