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, perceptual 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 un-certainties 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 Multi-Human Tracking and Intention-Recognition, 2013 - present

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 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.

3 Past Projects

3.1 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^2NG-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 race-track 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.


3.2 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.
We develop 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 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.

Figure 2: MAXQ task graph for WrightEagle.

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

3.3 Hierarchical Online Planning for Large MDPs, 2010 - 2012

We develop 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 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.


3.4 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 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 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

¹http://wiki.robocup.org/wiki/Soccer_Simulation_League
## Research Statement

**Aijun Bai**  

August 24, 2014

<table>
<thead>
<tr>
<th>Competitions</th>
<th>Games</th>
<th>Points</th>
<th>Goals</th>
<th>Win</th>
<th>Draw</th>
<th>Lost</th>
<th>Avg. Points</th>
<th>Avg. Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>RoboCup 2007</td>
<td>14</td>
<td>34</td>
<td>125 : 9</td>
<td>11</td>
<td>1</td>
<td>2</td>
<td>2.42</td>
<td>8.92 : 0.64</td>
</tr>
<tr>
<td>RoboCup 2008</td>
<td>16</td>
<td>40</td>
<td>74 : 18</td>
<td>13</td>
<td>1</td>
<td>2</td>
<td>2.50</td>
<td>4.63 : 1.13</td>
</tr>
<tr>
<td>RoboCup 2009</td>
<td>14</td>
<td>76</td>
<td>81 : 17</td>
<td>12</td>
<td>0</td>
<td>2</td>
<td>2.57</td>
<td>5.79 : 1.21</td>
</tr>
<tr>
<td>RoboCup 2010</td>
<td>13</td>
<td>33</td>
<td>123 : 7</td>
<td>11</td>
<td>0</td>
<td>2</td>
<td>2.54</td>
<td>9.47 : 0.54</td>
</tr>
<tr>
<td>RoboCup 2011</td>
<td>12</td>
<td>36</td>
<td>151 : 3</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>3.00</td>
<td>12.6 : 0.25</td>
</tr>
<tr>
<td>RoboCup 2012</td>
<td>21</td>
<td>58</td>
<td>104 : 18</td>
<td>19</td>
<td>1</td>
<td>1</td>
<td>2.76</td>
<td>4.95 : 0.86</td>
</tr>
<tr>
<td>RoboCup 2013</td>
<td>19</td>
<td>53</td>
<td>104 : 9</td>
<td>17</td>
<td>2</td>
<td>0</td>
<td>2.79</td>
<td>5.47 : 0.47</td>
</tr>
</tbody>
</table>

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

### Historical Results of WrightEagle

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](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.


### 3.5 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/](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


Research Statement
Aijun Bai
August 24, 2014


