Aijun Bai

Postdoctoral Researcher at UC Berkeley (Computer Science | Artificial Intelligence | Decision-Theoretic Planning)

aijunbai@gmail.com

Summary

- Specialties: Artificial Intelligence and Computer Science.
- Interests: decision-theoretic planning (MDPs/POMDPs), machine learning and robotics.
- Looking for positions in Artificial Intelligence, Robotics and/or Software/Algorithm Engineering.

Experience

Postdoctoral Researcher at UC Berkeley
April 2015 - Present (1 year 11 months)
Hierarchical decision-making and reinforcement learning.

Research Scholar at Carnegie Mellon University - Computer Science Department
November 2013 - March 2015 (1 year 5 months)
Real-time multi-object tracking based on particle filtering over sets

PHD Student at University of Science and Technology of China - Computer Science Department
September 2009 - November 2014 (5 years 3 months)
Markov models based planning and sensing under uncertainty

Researcher/Team Leader at WrightEagle Soccer Simulation Team - USTC
May 2007 - November 2013 (6 years 7 months)
Online planning for large markov decision processes with hierarchical decomposition

Undergraduate Student at University of Science and Technology of China - Computer Science Department
September 2005 - June 2009 (3 years 10 months)

Education

University of Science and Technology of China
Ph. D., Computer Science, 2009 - 2014
Activities and Societies: Worked in WrightEagle RoboCup team
Carnegie Mellon University
Visiting Research Scholar, Computer Science, 2013 - 2014
Activities and Societies: Worked in CoBot research group
University of Science and Technology of China

Bachelor's degree, Computer Science, 2005 - 2009
Activities and Societies: Worked in WrightEagle RoboCup team

Skills & Expertise

C++
Machine Learning
Algorithms
Research
Java
C
Python
Artificial Intelligence
Linux
Markov Decision Processes
Hidden Markov Models
Computer Science
Multi-agent Systems
Planning under Uncertainty
ROS
Matlab
Reinforcement Learning
Robotics
Big Data
ODPS
Data Mining
LaTeX
Programming

Honors and Awards

Postdoctoral Fellowship of UC Berkeley
UC Berkeley
September 2014
Berkeley, California, United States

Early Researcher Support of ICAPS
ICAPS Foundation
July 2014
Portsmouth, New Hampshire, United States

Travel Award of NIPS Foundation
NIPS Foundation
December 2013
Lake Tahoe, Nevada, United States

Ali Star of Alibaba Inc
Alibaba Inc

October 2013

Hangzhou, China

Scholarship of China Scholarship Council

China Scholarship Council

October 2013

World Champion of RoboCup Soccer Simulation 2D Competition

RoboCup Federation

July 2013

Eindhoven, The Netherlands

Glarun Scholarship of CETC

CETC

2013

Champion of RoboCup China Open Soccer Simulation 2D Competition

RoboCup China Open

December 2012

Hefei, China

Kwang-Hua Scholarship of USTC

University of Science and Technology of China

2012

Aegon-Industrial Responsibility Scholarship of USTC

University of Science and Technology of China

2012

Best Paper Nominee

RoboCup 2012 International Symposium

June 2012

Towards a Principled Solution to Simulated Robot Soccer, Aijun Bai, Feng Wu, and Xiaoping Chen, RoboCup-2012: Robot Soccer World Cup XVI, Lecture Notes in Artificial Intelligence, Vol. 7500, Springer Verlag, Berlin, 2013.

First Place of RoboCup Soccer Simulation 2D Free Challenge

RoboCup Federation

June 2012

Mexico City, Mexico

Second Place of RoboCup Soccer Simulation 2D Competition

RoboCup Federation

June 2012

Mexico City, Mexico

Champion of RoboCup China Open Soccer Simulation 2D Competition

RoboCup China Open

August 2011

Lanzhou, China

World Champion of RoboCup Soccer Simulation 2D Competition

RoboCup Federation

July 2011

Istanbul, Turkey

Champion of RoboCup China Open Soccer Simulation 2D Competition

RoboCup China Open

July 2010

Ordos, China

Second Place of RoboCup Soccer Simulation 2D Competition

RoboCup Federation

June 2010

Singapore, Singapore

Champion of RoboCup China Open Soccer Simulation 2D Competition

RoboCup China Open

November 2009

Dalian, China

World Champion of RoboCup Soccer Simulation 2D Competition

RoboCup Federation

June 2009

Graz, Austria

Second Place of RoboCup China Open Soccer Simulation 2D Competition

RoboCup China Open

December 2008

Zhongshan, China

Second Place of RoboCup Soccer Simulation 2D Competition

RoboCup Federation

July 2008

Suzhou, China

Champion of RoboCup China Open Soccer Simulation 2D Competition

RoboCup China Open

October 2007

Jinan, China

Outstanding Student Scholarship of USTC

University of Science and Technology of China

2008

Outstanding Student Scholarship of USTC

University of Science and Technology of China

2007

Outstanding Student Scholarship of USTC

University of Science and Technology of China
2006

Outstanding Freshman Scholarship of USTC
University of Science and Technology of China
2005

Publications

Markovian State and Action Abstractions for MDPs via Hierarchical MCTS
International Joint Conference on Artificial Intelligence (IJCAI), 2016 April 2016
Authors: Aijun Bai, Stuart Russell, Siddharth Srivastava

State abstraction is an important technique for scaling MDP algorithms. As is well known, however, it introduces difficulties due to the non-Markovian nature of state-abstracted models. Whereas prior approaches rely upon ad hoc fixes for this issue, we propose instead to view the state-abstracted model as a POMDP and show that we can thereby take advantage of state abstraction without sacrificing the Markov property. We further exploit the hierarchical structure introduced by state abstraction by extending the theory of options to a POMDP setting. In this context we propose a hierarchical Monte Carlo tree search algorithm and show that it converges to a recursively optimal hierarchical policy. Both theoretical and empirical results suggest that abstracting an MDP into a POMDP yields a scalable solution approach.

Online planning for large markov decision processes with hierarchical decomposition
ACM Transactions on Intelligent Systems and Technology (ACM TIST) July 2015
Authors: Aijun Bai, Feng Wu, Xiaoping Chen

Markov decision processes (MDPs) provide a rich framework for planning under uncertainty. However, exactly solving a large MDP is usually intractable due to the “curse of dimensionality”— the state space grows exponentially with the number of state variables. Online algorithms tackle this problem by avoiding computing a policy for the entire state space. On the other hand, since online algorithm has to find a near-optimal action online in almost real time, the computation time is often very limited. In the context of reinforcement learning, MAXQ is a value function decomposition method that exploits the underlying structure of the original MDP and decomposes it into a combination of smaller subproblems arranged over a task hierarchy. In this article, we present MAXQ-OP—a novel online planning algorithm for large MDPs that utilizes MAXQ hierarchical decomposition in online settings. Compared to traditional online planning algorithms, MAXQ-OP is able to reach much more deeper states in the search tree with relatively less computation time by exploiting MAXQ hierarchical decomposition online. We empirically evaluate our algorithm in the standard Taxi domain—a common benchmark for MDPs—to show the effectiveness of our approach. We have also conducted a long-term case study in a highly complex simulated soccer domain

and developed a team named WrightEagle that has won five world champions and five runners-up in the recent 10 years of RoboCup Soccer Simulation 2D annual competitions. The results in the RoboCup domain confirm the scalability of MAXQ-OP to very large domains.

Please: Palm leaf search for pomdps with large observation spaces (abstract)

Twenty-Fifth International Conference on Automated Planning and Scheduling (ICAPS 2015) June 2015

Authors: Zongzhang Zhang, David Hsu, Wee Sun Lee, Zhan Wei Lim, Aijun Bai

Trial-based asynchronous value iteration algorithms for large Partially Observable Markov Decision Processes (POMDPs), such as HSVI2, FSVI and SARSOP, have made impressive progress in the past decade. In the forward exploration phase of these algorithms, only the outcome that has the highest potential impact is searched. This paper provides a novel approach, called Palm LEAf Earch (PLEASE), which allows the selection of more than one outcome when their potential impacts are close to the highest one. Compared with existing trial-based algorithms, PLEASE can save considerable time to propagate the bound improvements of beliefs in deep levels of the search tree to the root belief because of fewer point-based value backups. Experiments show that PLEASE scales up SARSOP, one of the fastest algorithms, by orders of magnitude on some

POMDP tasks with large observation spaces.

Intention-Aware Multi-Human Tracking for Human-Robot Interaction via Particle Filtering over Sets

AAAI 2014 Fall Symposium: AI for Human-Robot Interaction (AI-HRI 2014) November 2014

Authors: Aijun Bai, Reid Simmons, Manuela Veloso, Xiaoping Chen

We present a novel particle filtering over sets (PFS) approach to the intention-aware multi-human tracking problem in the domain of human-robot interaction. From a multiobject tracking point of view, our approach avoids directly performing observation-to-target association by using a set formulation. The experiments in PETS2009 dataset show that the overall tracking performance is robust. The real robot experiments indicate that a robot integrated with this approach is able to track humans, and understand their intentions in terms of moving and staying. In future work, we plan to apply PFS in complex social tasks with multiple intentions.

Thompson Sampling based Monte-Carlo Planning in POMDPs

24th International Conference on Automated Planning and Scheduling (ICAPS 2014) June 2014

Authors: Aijun Bai, Feng Wu, Zongzhang Zhang, Xiaoping Chen

Monte-Carlo tree search (MCTS) has been drawing great interest in recent years for planning under uncertainty. One of the key challenges is the trade-off between exploration and exploitation. To address this, we introduce a novel online planning algorithm for large POMDPs using Thompson sampling based MCTS that balances between cumulative and simple regrets. The proposed algorithm Dirichlet-Dirichlet-

NormalGamma based Partially Observable Monte-Carlo Planning (D2NG-POMCP) treats the accumulated reward of performing an action from a belief state in the MCTS search tree as a random variable following an unknown distribution with hidden parameters. Bayesian method is used to model and infer the posterior distribution of these parameters by choosing the conjugate prior in the form of a combination of two Dirichlet and one NormalGamma distributions. Thompson sampling is exploited to guide the action selection in the search tree. Experimental results confirmed that our algorithm outperforms the state-of-the-art approaches on several common benchmark problems.

Bayesian Mixture Modelling and Inference based Thompson Sampling in Monte-Carlo Tree Search

Advances in Neural Information Processing Systems 26 (NIPS 2013) December 2013

Authors: Aijun Bai, Feng Wu, Xiaoping Chen

Monte-Carlo tree search (MCTS) has been drawing great interest in recent years for planning and learning under uncertainty. One of the key challenges is the trade-off between exploration and exploitation. To address this, we present a novel approach for MCTS using Bayesian mixture modeling and inference based Thompson sampling and apply it to the problem of online planning in MDPs. Our algorithm, named Dirichlet-NormalGamma MCTS (DNG-MCTS), models the uncertainty of the accumulated reward for actions in the search tree as a mixture of Normal distributions. We perform inferences on the mixture in Bayesian settings by choosing conjugate priors in the form of combinations of Dirichlet and NormalGamma distributions and select the best action at each decision node using Thompson sampling. Experimental results confirm that our algorithm advances

the state-of-the-art UCT approach with better values on several benchmark problems.

Towards a Principled Solution to Simulated Robot Soccer

RoboCup-2012: Robot Soccer World Cup XVI, Lecture Notes in Artificial Intelligence 2013

Authors: Aijun Bai, Feng Wu, Xiaoping Chen

The RoboCup soccer simulation 2D domain is a very large testbed for the research of planning and machine learning. It has competed in the annual world championship tournaments in the past 15 years. However it is still unclear that whether more principled techniques such as decision-theoretic planning take an important role in the success for a RoboCup 2D team. In this paper, we present a novel approach based on MAXQ-

OP to automated planning in the RoboCup 2D domain. It combines the benefits of a general hierarchical structure based on MAXQ value function decomposition with the power of heuristic and approximate techniques. The proposed framework provides a principled solution to programming autonomous agents in large stochastic domains. The MAXQ-OP framework has been implemented in our RoboCup 2D team, WrightEagle. The empirical results indicated that the agents developed with this framework and related techniques reached outstanding performances, showing its potential of scalability to very large domains.

Online Planning for Large MDPs with MAXQ Decomposition

11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)

Authors: Aijun Bai

Markov decision processes (MDPs) provide an expressive framework for planning in stochastic domains. However, exactly solving a large MDP is often intractable due to the curse of dimensionality. Online algorithms help overcome the high computational complexity by avoiding computing a policy for each possible state. Hierarchical decomposition is another promising way to help scale MDP algorithms up to large domains by exploiting their underlying structure. In this paper, we present an effort on combining the benefits of a general hierarchical structure based on MAXQ value function decomposition with the power of heuristic and approximate techniques for developing an online planning framework, called MAXQ-OP. The proposed framework provides a principled approach for programming autonomous agents in a large stochastic domain. We have been conducting a long-term case-study with the RoboCup soccer simulation 2D domain, which is extremely larger than domains usually studied in 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.

WrightEagle and UT Austin Villa: RoboCup 2011 Simulation League Champions

Authors: Aijun Bai, Xiaoping Chen, Patrick MacAlpine, Daniel Urieli, Samuel Barrett, Peter Stone

The RoboCup simulation league is traditionally the league with the largest number of teams participating, both at the international competitions and worldwide. 2011 was no exception, with a total of 39 teams entering the 2D and 3D simulation competitions. This paper prese- nts the champions of the competitions, WrightEagle from the University of Science and Technology of China in the 2D competition, and UT Austin Villa from the University of Texas at Austin in the 3D competition.

Projects

Markovian State and Action Abstractions for MDPs

November 2015 to Present

Members:Aijun Bai

State abstraction is an important technique for scaling MDP algorithms. As is well known, however, it introduces difficulties due to the non-Markovian nature of state-abstracted models. Whereas prior approaches rely upon ad hoc fixes for this issue, we propose instead to view the state-abstracted model as a POMDP and show that we can thereby take advantage of state abstraction without sacrificing the Markov property. We further exploit the hierarchical structure introduced by state abstraction by extending the theory of options to a POMDP setting. In this context we propose a hierarchical Monte Carlo tree search algorithm and show that it converges to a recursively optimal hierarchical policy. Both theoretical and empirical results suggest that abstracting an MDP into a POMDP yields a scalable solution approach.

Meta-Reasoning Approach for Real-Time Online Planning

April 2015 to Present

Members:Aijun Bai

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.

Semi-Markov Process based User Behavior Modelling

December 2014 to March 2015

Members:Aijun Bai

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.

Multi-Human Tracking and Intention-Recognition

November 2013 to October 2014

Members:Aijun Bai

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. 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. 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 in terms of CLEAR MOT metrics. 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 \ref{fig:benchmark} 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.

Bayesian Monte-Carlo Tree Search with Thompson Sampling

2012 to 2013

Members:Aijun Bai

Monte-Carlo tree search (MCTS) has been drawing great interest recently in domains of planning and learning under uncertainty. 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, DNG-MCTS produces competitive results in the CTP domain, and converges faster in the domains of racetrack and sailing with respect to sample complexity. Experimental results in POMDPs show that D ^{2}NG-POMCP outperforms the state-of-the-art algorithms in both RockSample and PocMan domains. Related work is published in NIPS 2013 and ICAPS 2014.

An Intelligent Service System with Multiple Robots

2013 to Present

Members:Aijun Bai

A multi-robot project where an intelligent bartender-waiter robot system serves people in a bar. In this system, KeJia robot 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.

Hierarchical Online Planning for Large MDPs

2010 to 2012

Members:Aijun Bai

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 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. We have also been conducting a long-term case-study with the RoboCup soccer simulation 2D domain, 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. Some of the results have been published in {\bf RoboCup Symposium} 2011, AAMAS 2012, ARMS 2012 and RoboCup Symposium 2012.

WrightEagle Soccer Simulation 2D Team

2007 to 2013

Members:Aijun Bai

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. 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. 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 \emph{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.

Reversi Game

2006 to 2008

Members:Aijun Bai

I developed 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.

Volunteer Experience

Programme Committee at International Joint Conferences on Artificial Intelligence (IJCAI)
2017 - Present

Programme Committee at Association for the Advancement of Artificial Intelligence (AAAI)
2016 - Present

Programme Committee at International Joint Conferences on Artificial Intelligence (IJCAI)
2015 - Present

Programme Committee at International Joint Conferences on Artificial Intelligence (IJCAI)
2016 - Present

Organizing Committee at RoboCup Federation
2012 - 2013
RoboCup Soccer Simulation 2D League

Technical Committee at RoboCup Federation
2011 - 2012
RoboCup Soccer Simulation 2D League

Organizations

Association for the Advancement of Artificial Intelligence (AAAI)
Member
2013 to 2014

Languages

English (Full professional proficiency)
Chinese (Native or bilingual proficiency)

Aijun Bai

Postdoctoral Researcher at UC Berkeley (Computer Science | Artificial Intelligence | Decision-Theoretic Planning)

aijunbai@gmail.com

Contact Aijun on LinkedIn