International Journal of Multimedia and Ubiquitous Engineering
Vol. 9, No. 10 (2014), pp. 23-32
http://dx.doi.org/10.14257/ijmue.2014.9.10.03

Using Modified UCT Algorithm Basing on Risk Estimation Methods in Imperfect Information Games

Jiajia Zhang[1] and Xuan Wang[1]

[1]*Intelligence Computing Research Center Harbin Institute of Technology Shenzhen Graduate School C302, HIT Campus Shenzhen University Town, NanShan District, XiLi, Shenzhen 518055, P. R. China*
jiachina@sohu.com

Abstract

Risk dominance and payoff dominance strategy are two complementary parts of the game theory decision strategy. While payoff dominance is still the basic principle in perfect information, two player games, risk dominance has shown its advantages in imperfect information conditions. In this paper, we first review the related work in the area of estimation methods and the influence of risk factors on computing game equilibrium. Then a new algorithm, UCT-Risk is proposed in this paper, which is a modification of UCT (UCB apply to Trees) algorithm based on risk estimation methods. Finally, we implement the proposed algorithm in SiGuo game, a popular imperfect information game in China. The experimental result of the new algorithm shows it correctness and effectiveness.

Keywords: *imperfect information games; risk dominance, UCT*

1. Introduction

Games can be classified as perfect or imperfect information conditions, which are based on whether or not players have the whole information of the game [1]. Researches on perfect information games have achieved some kind of success. For example, "Deep Fritz" [2], one of the best Computer-Chess players, is at the level of top professional human players. In imperfect information games such as poker, certain relevant details are withheld from the players. Many researches regarding the imperfect information game systems based on Monte-Carlo approaches [1] and alpha-beta search have been conducted. The bridge and heart game of Alberta University [3] is a good example in this field. However, when the branch sizes of the game tree become huge, these approaches cannot perform efficiently. Thus, the multiple equilibrium theory and methods are studied recent years as another approach to solve imperfect information games [4].

Different standpoints are proposed by many researchers about the analysis and judgment of multiple equilibrium. Using Nash equilibrium as the result of strategy selection is not always consequent. Risk dominance and payoff dominance are two related refinements of the Nash equilibrium (NE) solution concept in game theory, defined by John Harsanyi and Reinhard Selten in 1988 [5]. The research of Cooper consists that risk dominance supposes correct approach of dealing with imperfect information conditions [6]. The research on risk factors has been an important part of the theory of game equilibrium selection. The imperfect information conditions lead the problem of multiple equilibrium selection and risk dominance theory and related methods is playing an important roll in this field.

There are two main contributions of this paper. The role of risk estimation methods for search in large imperfect-information game trees is examined. And then, UCT algorithm, which is significantly increasing the playing strength of several domains in our research, is

ISSN: 1975-0080 IJMUE

modified basing on risk estimation methods. The new algorithm, called "UCT-Risk" in this paper, is tested for its performance in SiGuo game.

This paper is organized as follows. Section 2 briefly introduces related work about the traditional algorithms and UCT algorithm. Section 3 shows our research about risk estimation methods in imperfect information conditions. Section 4 introduces the modified UCT algorithm, "UCT-Risk" and its performance in practice. And finally, Section 5 gives the conclusions.

2. Monte Carlo Method and UCT

Imperfect information game has been one of the most important problems in computer games. The ideal solution, at least for two-player zero-sum games, is to use a solution technique that can produce Nash equilibrium. However, this is usually computationally infeasible in most imperfect information domains for its precondition about guaranteeing perfect play against perfect opponents.

One popular way of dealing with imperfect information has been to avoid the issue. Instead of solving a full game, perfect information worlds from the game are sampled and solved either exactly or heuristically [3]. This method is Monte-Carlo sampling, which is applied to the imperfect information game problems by Bampton [1] firstly in 1994, has grasped much more attention in recent years. Monte-Carlo sampling, also called computer random simulation method, deals with imperfect information problem by a heuristics search method. The main idea is to create a subset of possible conditions by random sampling and to find out the best solution for them. Based on a statistical hypothesis that the optimal solution for a random-sampling subset is comparable to the global optimal solution, the solution is then treated as acceptable for the whole problem.

While Monte-Carlo method greatly increases the scale of possible worlds, basing on custom game tree search methods, imperfect information games are also difficult for computers to play well. Go is a typical example of this. As a new raising method, UCT works particularly well in Go in last few years.

UCT is the extension of UCB1 to mini-max tree search. The idea is to consider each node as an independent bandit, with its child nodes as independent arms. Instead of dealing with each node once iteratively, it plays sequences of bandits within limited time, each beginning is from the root of v the game tree and ending is at one leaf.

The main process of UCT algorithm works as explained below and also depicted in Figure 1. In the process, the word "value" means the calculated result by selected UCT policy.

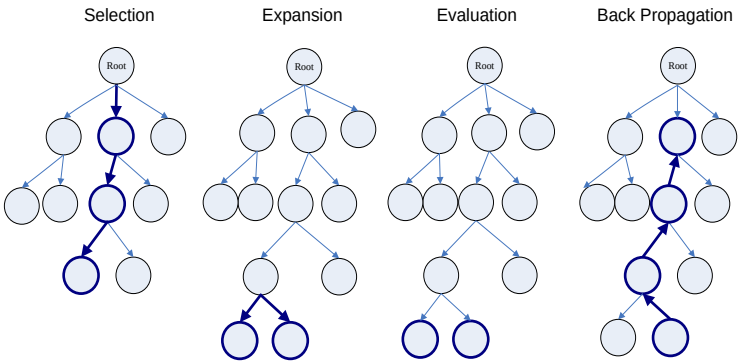

Figure 1. The Process of UCT Algorithm Operation for One Loop

UCT outperforms alpha-beta search [7] for at least three major advantages. First, it works in an anytime manner. The algorithm can be stopped at any moment, and its performance can be somehow good and this is not the case of alpha-beta search. Second, UCT is robust as it automatically handles uncertainty in a smooth way. At each node, the computed value is the mean of the values because each child weighted by the frequency of visits. Third, the tree grows in an asymmetric manner and it explores the good moves more deeply. What is more, this is achieved in an automatic manner.

3. Estimating Risks in Imperfect Information Conditions

Risk dominance and payoff dominance are two related refinements of the Nash equilibrium (NE) solution concept in game theory, defined by John Harsanyi and Reinhard Selten. Their research proposes that payoff dominance is reasonable when a NE has Pareto superior compare to other equilibriums. All of the player will follow payoff dominance because it provides best payoff expectation than others. However, when the game environment of the players' information is imperfect, risk dominance can usually performs more reasonable equilibrium. In Paul Straub's research [8], basing on the experiment on human under imperfect information conditions, the similar conclusion is confirmed.

Imperfect information domains, like Texas Hold'em poker and Siguo game, has very similar proprieties like the examinations of upper researchers. Figure 2 shows a normal search tree of perfect information game. Evaluation function evaluates the leaves and the best branch is selected to trace the strategy array.

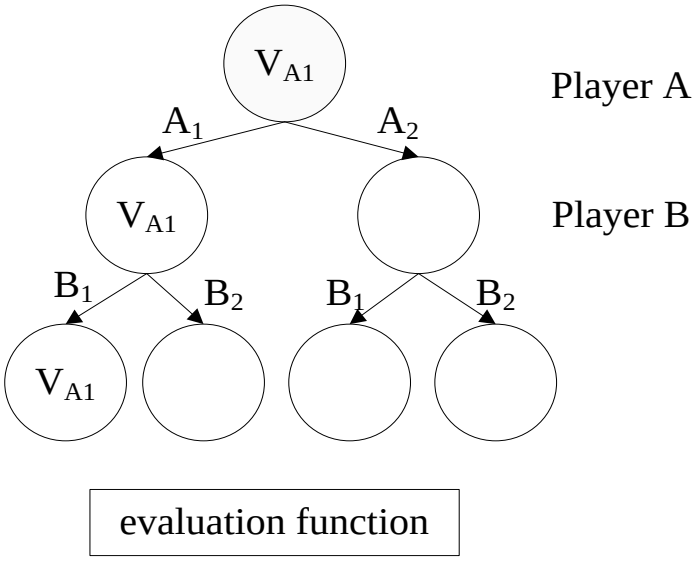

Figure 2. Normal Search Tree of Computer Game

In imperfect information games, Perfect information Monte Carlo method (PIMC) is widely used in upper process [9]. Figure 3 shows the process of PIMC in which I means an imperfect information state of the game, W means the set of possible real worlds of the state, S means the sample set of W and M means the set of best strategies of each world in S. the process of choosing mi from Si is same to the Figure 2 shows.

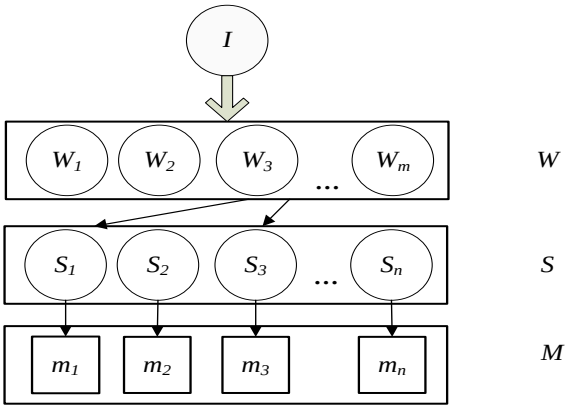

Figure 3. Search Tree of Imperfect Information Game

In imperfect information conditions, there are at least two kinds of possible deviations between expect and exact profit of the selected strategy. Firstly, the evaluation function can't provide an accurate value of profit because the sample is exactly a sub-set of the whole worlds. Secondly, the strategies that opponents applied may differ from what are expected. The two kinds of deviations are donated as L_{wI} and L_{mII} in this paper. Furthermore, the risk lost is defined as L_{wm} and calculated as formula 1 and Figure 4 gives a further explanation.

$$L_{wm} = L_{wI} + L_{mII} \tag{1}$$

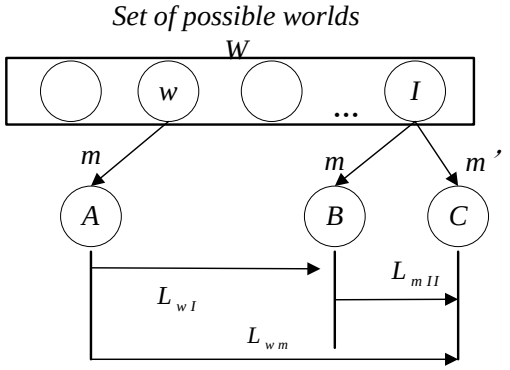

Figure 4. Further Explanation of the Calculation of Risk Lost

$$L_{wI} = E_w^m - E_I^m \tag{2}$$

$$L_{mII} = E_I^m - E_I^{m'} \tag{3}$$

Formula 2 and 3 shows the exact calculate methods of risk lost. E_w^m means the expect profit of strategy array m in world w and E_I^m means the exact profit of strategy m. E_I^m means

the expect profit when strategy array m is adopted and $E_I^{m'}$ means the exact profit basing on real strategy array m'.

Formula 1 gives calculation of the strategy array m's risk lost under world w. The synthesized estimation of m's risk lost in all possible worlds set W can be calculated as formula in which n means the number of possible worlds. The details of the donations and the property of risk lost in different information conditions can be informed in our former research [10].

$$L_{Wm} = \frac{1}{n}\sqrt{\sum_{i=1}^{n} L_{w_i m}{}^2} = \frac{1}{n}\sqrt{\sum_{i=1}^{n} (L_{w_i I} + L_{mII})^2} \tag{4}$$

Formula 4 provides an idealized approach to calculate the risk lost of the candidate strategy arrays. However, it can't be applied to strategy decision methods directly in imperfect information conditions. That is because E_I^m and $E_I^{m'}$ cannot be exactly calculated in the decision state of imperfect information games. Thus, a method that can calculate the approximate result of risk lost value in imperfect information conditions which is basing on the set of sample worlds is provides as following.

Define: $\overline{E_s}$ is the average profit of sample set S and can be calculated as following:

$$\overline{E_s} = \frac{1}{tk}\sum_{i=1}^{t}\sum_{j=1}^{k} E_i^j (i \in S, j \in M) \tag{5}$$

Basing on formula (5), the method that can calculate the approximate result of risk lost value of strategy array m in imperfect information conditions can be defined as following:

$$\begin{aligned} L_{Wm} &= \frac{1}{n}\sqrt{\sum_{i=1}^{n} L_{w_i m}{}^2} = \frac{1}{n}\sqrt{\sum_{i=1}^{n} (L_{w_i I} + L_{mII})^2} \\ &= \frac{1}{n}\sqrt{\sum_{i=1}^{n} (E_{w_i}^m - E_I^m + E_I^m - E_I^{m'})^2} \\ &= \frac{1}{n}\sqrt{\sum_{i=1}^{n} (E_{w_i}^m - E_I^{m'})^2} \\ &\approx \frac{1}{t}\sqrt{\sum_{i=1}^{t} (E_{w_i}^m - \overline{E_s})^2} \qquad \text{in witch}(w_i \in S) \end{aligned} \tag{6}$$

The position that connected by approximately equal signal is the process of approximate calculation by using $\overline{E_s}$ and sample set S.

Basing on upper methods, the agent can estimate the risk lost of each candidate strategy arrays in strategy decision. Considered the advantages of risk dominance strategy in imperfect information problems, a new strategy decision algorithm is provided in this paper which is a combination of UCT algorithm and risk lost estimation methods.

4. UCT-Risk Algorithm

UCT algorithm is based from muti-armed bandit problem (UCB). Auer and Al. provided several deterministic policies, which are called UCB1, UCB2, and UCB-normal [11]. In this paper, algorithm UCB1 is introduced as the foundation of this chapter.

Algorithm 1: Deterministic policy: UCB1

Initialization: Play each machine once.

$$\bar{x}_j + \sqrt{\frac{2 \ln n}{n_j}} \tag{7}$$

Loop: Play machine j that maximizes (7), where $\bar{x}_j$ is the average profit obtained form strategy j. nj is the number of times strategy j has been checked so far, and n is the overall number of plays done so far.

Conclusion: the strategy j that maximizes (7) is chosen as the final result of the search process.

Other types of UCT policies have the similar mode as UCB1 all of which has an appendix part behind the average profit. Basing on our research on the performance of UCT algorithm in imperfect information games, the appendix part gives UCT algorithm a primer character of risk avoidance which is also the core point of this paper.

In this sense, algorithm UCT-Risk can be defined as following:

Algorithm: Deterministic policy: UCB-Risk

· Initialization: Play each branch once.

$$\bar{x}_j - R \frac{1}{t} \sqrt{\sum_{i=1}^{t} (E_{w_i}^{m} - \overline{E_s})^2} \tag{8}$$

· Loop: Play machine j that maximizes (8), where $\bar{x}_j$ is the average profit obtained form strategy j. R is the influence factor of risk dominance and the appendix part is the calculation formula of estimated risk lost of strategy j.

Conclusion: the strategy j that maximizes (8) is chosen as the final result of the search process.

Algorithm UCT-Risk uses the calculation formula of estimated risk lost as the appendix part of average profit. Parameter R is used to control the influence of risk dominance. When R=0, algorithm UCT-Risk changes to ε-GREEDY policy of UCT and when R =1, UCT-Risk is following the pure risk dominance equilibrium.

The core point of UCT-Risk algorithm is the factor of risk lost is adopted as the appendix of the strategy's average profit. Both payoff dominance and risk dominance factors are considered while the influence of them can be adjusted.

5. Experiments and Performance Evaluation

5.1. Parameters Set in Experiments

In this section, an imperfect information game called SiGuo game is chosen as the experiments domain. This game is known as one of the most complex imperfect information games as its huge scale of possible words which also makes it as a good test platform of UCT algorithm. The details of the game rules are described in [12]. Basing our research, an

analysis [13] of the complexity about several popular games is shown in Table 1. In this table, Siguo game has a much larger number of initial possible worlds than Go and Poker games.

Table 1. Analysis on the Complexity of Several Games

	Number of initial worlds	Average number of worlds in a round	information decreases
Go	19*19	10*19	Slow
Poker	$8.45*10^{16}$	$1.72*10^{7}$	Fast
Siguo Game	$3.57*10^{53}$	$1.10*10^{26}$	Normal

In our experiments, UCB-1, UCB-Risk and ε-GREEDY policy are tested as strategy decision policy. Parameter R is tested as the performance of risk dominance strategy in imperfect information games.

5.2. Experiments against Other UCT Policies

In our experiments, the agent basing on UCT-Risk algorithm is tested with UCT-Turned and ε-GREEDY agents as its opponents. The Siguo game system with UCT agent has been exhibited in 2012 2nd IEEE International Conference on Cloud Computing and Intelligence Systems. Parameter R is set from 0 to 1 and each set of 0.2 values is played 20 rounds. For clear exhibition, the exact evaluate value of the game is mapped to the range of -100~100.

Table 2. Performance of UCT-Risk against UCT-Turned

R	Win	Lose	Average profit	Max lost
0	9	91	-77.23	31.55
0.2	35	65	-24.91	30.14
0.4	58	42	19.41	15.39
0.6	52	48	4.10	14.09
0.8	46	54	-9.28	10.02
1.0	40	60	-17.11	10.98

Table 3. Performance of UCT-Risk against ε-GREEDY

R	Win	Lose	Average profit	Max lost
0	44	56	-4.40	29.68
0.2	49	51	7.98	26.65
0.4	61	39	22.51	18.20
0.6	66	34	20.69	18.03
0.8	71	29	21.37	13.32
1.0	58	42	15.06	9.25

Table 2 and Table 3 show the performance of UCT-Risk algorithm with parameter R against UCT-Turned and UCT-ε-GREEDY.

When comparing UCT-Turned policy, given 20 samples each value of R, UCT-Risk algorithm wins 58% of the game when R is set 0.4 and 52% when R is 0.6. However, as one of the best previous approaches of imperfect information problem, UCT-Turned performs better then UCT-Risk with other values of R. However, UCT-Risk algorithm takes obvious advantages when against UCT-ε-GREEDY policy, which is the donation of payoff dominance approach.

The last line of table 2 and 3 shows the average max lost of the agent's profits in games which are calculated by the max fluctuation between neighboring steps. In this sense, the contribution of the UCT-Risk's appendix of risk lost is obvious. With the increasing value of R, the influence of risk lost appendix effectively decrease the fluctuation of the profits between neighboring steps. That means the agents adopts UCT-Risk algorithm becomes more cautious in imperfect information conditions and shows more capacity of risk prevention.

For testing the contribution of risk lost appendix of UCT-Risk, the variance of agent's profits are observed as the critical factor in the following experiments.

Table 4. The Variance of Agent's Profits basing on Different Algorithms

Steps	UCT-Turend	ε-GREEDY	UCT-Risk R=0.3	UCT-Risk R=0.7	UCT-Risk R=1.0
16	56.25	1209	1	16	49
32	240.25	90.25	6.25	156.25	16
48	102.25	200	42.25	121	36
64	272.25	100	110.25	276	240.25
80	882.25	200.25	9	169	81
96	216	230.25	484	2.25	0.25
112	225	264	400	220.25	156.25
128	144	420.25	361	9	25

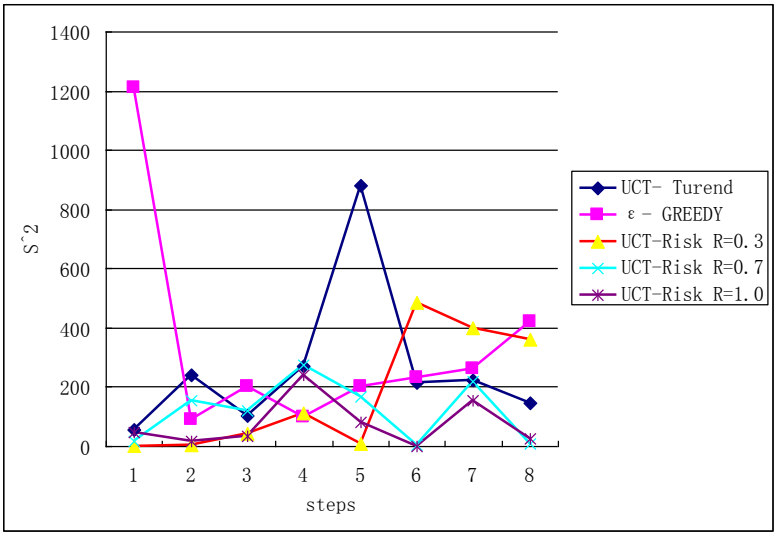

Figure 5. The Variance of Agent's Profits basing on Different Algorithms

Table 4 shows the experiments of variance of agent's profits basing on different algorithms and Figure 5 illustrates the same contents with figure. The whole rounds of game are separated every 16 steps to calculate the profits' variance. It can be observed that UCT-Risk's variance keeps at low fields comparing with UCT-Turned and ε-GREEDY policy. Also, the larger parameter R is set, the lower the variance of profits keeps. The results of this experiment also establish that UCT-Risk can effectively alleviate the fluctuation of agents' profit and the contribution of risk lost prevention can be concluded.

6. Conclusions

In this paper, the definition and estimation method of risk lost in imperfect information conditions is introduced. And a new algorithm, named UCT-Risk is proposed which is based on the combination of UCT algorithm and risk lost estimation methods. Using risk lost formula as an appendix and parameter R as its influence controller, the UCT-Risk algorithm considers the balance of payoff dominance and risk dominance in imperfect information conditions. To evaluate the performance of the new algorithm, Siguo game which is a typical imperfect information game in China, is introduced and used as the examination platform. In over 200 rounds of games, UCT-Risk algorithm performs better than ε-GREEDY algorithm and UCT-Turned algorithm. In the performs of different policies and parameter sets of UCT algorithm have been discussed in this paper and further tested by SiGuo game system. In over 200 rounds of games, UCT-Risk algorithm shows obvious advantages to ε-GREEDY policy and even a little better performed than UCT-Turned when parameter R is suitable set (0.4~0.6). A further experiment is developed to examine the variance of profits in the process of a round. The result of the experiments confirmed the contribution of UCT-Risk algorithm on risk prevention further. Basing on the research of this paper, at least two points can be concluded. Firstly, besides payoff dominance, risk dominance ideology has its advantages on dealing with imperfect information conditions. Secondly, UCT-Risk, with well adjustment of the influence parameter of risk lost, can perform as well or better as the best previous approaches.

Acknowledgements

We would like to thank Xiao Ma for the theory direction for our system. We also appreciate Jing Lin for his fundamental work of the Siguo system.

References

[1] H. J. Bampton, "Solving imperfect information games using the Monte Carlo heuristic", Master thesis, University of Tennessee, Knoxville, (**1994**).
[2] D. Papp, "Dealing with imperfect information in Poker", MSC. Thesis, University of Alberta, (**1998**), pp. 1-2.
[3] J. Long, N. R. Sturtevant and M. Buro, "Understanding the success of perfect information Monte Carlo sampling in game tree search", Association for the Advancement of Artificial Intelligence AAAI-10, vol. 1, (**2010**), pp. 134-140.
[4] N. R. Sturtevant, "Current challenges in multi-player game search", Proceedings of the 4th International Conference on Computers and Games, Ramat-Gan, Israel, (**2004**), pp. 285–300.
[5] J. C. Harsanyi, and R. Selten, "A general theory of equilibrium selection in games", Discrete Applied Mathematics, vol. 26, no. 1, (**1990**), pp. 126-127.
[6] D. Lee, "Managing Presence Information for Online 2D Games", International Journal of Advanced Science and Technology, vol. 34, no. 2, (**2011**) September, pp. 113—118.
[7] D. Koller and N. Megiddo, "The complexity of two-person zero-sum games in extensive form", Game and Economic Behavior, vol. 4, no. 4, (**1992**), 10, pp. 528-552.
[8] P. G. Straub, "Risk dominance and coordination failures in static games", Quart. Rev. Econ. Finance 35, vol. 35, no. 4, (**1995**), pp. 339–363.
[9] J. Long, N. R. Sturtevant and M. Buro, "Understanding the success of perfect information Monte Carlo sampling in game tree search", Association for the Advancement of Artificial Intelligence AAAI-10, vol. 1, (**2010**), pp. 134-140.
[10] J. Zhang and X. Wang, "The research of risk analysis and estimate method for machine game", Journal of High Technology Letters, vol. 11, (**2013**), inpress.
[11] L. Kocsis and C. Szepesvari, "Bandit based Monte-Carlo Planning, 15th European Conference on Machine Learning (ECML)", (**2006**), pp. 282-293.
[12] X. Wang, J. Zhang and X. Xu, *et al.* "Risk dominance strategy in imperfect information multi-player military chess game", International Swaps and Derivatives Association, Gaoxiong, China, (**2008**), pp. 596-601.

[13] X. Ma, W. Xuan and W. Xiaolong, "The Information Model for a Class of Imperfect Information Game", Journal of Computer Research and Development, vol. 47, no. 12, (**2010**), pp. 2100-2109.

Authors

Jiajia Zhang, he was born in 1984. Received the B.A's and M.A's. degrees in computer science from Harbin Institute of Technology, Harbin, China, in 2006 and 2009 respectively. Since 2009, he has been a Ph.D. degree candidate in computer science from ShenZhen graduate school of Harbin Institute of Technology, ShenZhen, China. His current research interests include artificial intelligence and computer game.

Xuan Wang, he was born in 1969. He has been professor of Harbin institute of Technology science 2006. His main research interests are artificial intelligence, computer network security and computational linguistics.

