International Journal of u- and e- Service, Science and Technology
Vol. 7, No. 6 (2014), pp. 33-46
http://dx.doi.org/10.14257/ijunesst.2014.7.6.04

The Naive Bayesian Algorithm-based Prisoner's Dilemma Game Model

Xiuqin Deng[1] and Jiadi Deng[2]

[1]*School of Applied Mathematics, Guangdong University of Technology, Guangzhou City, P. R. China*
deng_xiuqin@126.com
[2]*Department of Computer Science and Technology, Tsinghua University, Beijing City, P. R. China*

Abstract

Prisoners' dilemma is a typical game theory issue. In this study, it was treated as an incomplete information game to establish a related machine learning model using a naive Bayesian classification method. The model established was referred to as the Bayes model. Using this model, the incomplete information game was soluble with the assistance of statistical machine learning. This study proceeded as follows: firstly, four typical models were run against the Bayes model some 10,000 times. The total incomes of the models recorded suggested that Bayes model was more advantageous than other models. Even in a multi-player prisoners' game, Bayes model also presented the desired level of performance and accrued a higher income than other models. Further statistical analysis implied that the Bayes model and the widely accepted optimum strategy tit-for-tat (TFT) model showed a tendency to be prone to defection. Secondly, according to the games run on the natural Bayes model, as well as the natural TFT model, it was found that the Bayes model accrued more benefits than the TFT model on average. Finally, comparison of the Bayes model with the TFT model revealed that the Bayes model was better. This demonstrated the efficacy of the Bayes model constructed in this study and moreover, provided a novel idea for solving the problem of an incomplete information game.

Keywords: *game, prisoners' dilemma, machine learning, Bayesian algorithm, incomplete information game*

1. Introduction

Incomplete information games are influenced by the private information owned by at least one game player, such as current game state, mechanism of other players in decision-making, the game state of other players, the reward/punishment mechanism of the game, *etc*. [1]. However, due to the absence of its optimal, or relatively optimal solution (at any certain game state), such an incomplete information game is insoluble by traditional methods. This is attributable to the fact that the strategy of incomplete information games is restrained by, and related to, the other players. Harsanyi analyzed an incomplete information game using Bayesian game player strategies [2] and provided the methods for modeling and analyzing game problems. Zinkevich investigated incomplete information games using a Nash equilibrium and minimum regret method [3] and proposed some game strategies for improving poker game problems.

ISSN: 2005-4246 IJUNESST

As a branch of artificial intelligence and a cutting-edge research topic, machine learning has been paid great attention in related fields in recent years. Machine learning is defined as a research method aiming at obtaining a more desired approximate solution through the general law yielded by the analysis of a large amount of data [4]. Statistical machine learning is a branch of machine learning. It integrates statistical theory into machine learning by combining probability theory and stochastic mathematical knowledge with machine learning to improve the efficiency and accuracy [5-6]. The Bayesian classification algorithm is a commonly used machine learning method [7]. The simplified model of naive Bayesian classification is often used in text classification.

The prisoners' dilemma game is a classic cooperation and selection problem based on the assumption of selfish human motives [8]. It is popular and widely applied in mathematics and economics [9]. For a long time, it has attracted great interest from mathematics and economics researchers around the world. Game theory was born in the mid-twentieth Century and was founded by von Neumann (a famous mathematician and founding father of computing) and Morgenstern (a famous economist). Game theory brought radical changes to economics and provided a standard analysis tool for economists. In light of the contributions of game theory to economics, the Royal Swedish Academy of Sciences awarded Nobel Prizes for economics to Nash, Harsanyi, and Selten in 1994 and Aumann and Schelling in 2005 respectively [10]. In the famous artificial intelligence algorithm competition of prisoners' dilemma, Axelrod concluded that a TFT model was the optimum solution through brute competition in participating algorithms [11]. Miller introduced an automaton model to simplify and analyze prisoners' dilemma [12] and proposed a more general prisoners' dilemma decision analysis method. He also applied the model to other generalization problems. Lin uses a genetic algorithm to infer the optimum solution to the prisoners' dilemma [13]. By analysis, he obtained a better solution than TFT using a traditional static decision algorithm.

In this study, prisoners' dilemma is treated as a game with incomplete information. It satisfies the conditions for an incomplete information game, namely, the players of each game are incapable of determining the choice of their rival in any current station. Subsequently, a naive Bayesian classification method is used to construct the machine learning model for prisoners' dilemma in an attempt to solving it through statistical machine learning.

2. Construction of the Model

2.1. Prisoners' Dilemma Model

In this game, cooperation or defection is determined by two individuals. If the two individuals are mutually cooperative, they both earn incomes *R*; if they defect to each other, the incomes of both sides are *P*; if one individual is cooperative while the other is in a state of betrayal, the cooperative one gains *S*, while the treacherous player gains *T* (Table 1). Here, $T>R>P>S$, and $2R>S+T$. The latter formula means that total income of the two cooperative individuals is always larger than that gained in case of one individual's treachery. However, with regard to individuals, the incomes earned by defection to cooperation are greater than that by cooperation to cooperation.

Table 1. The Game Information

(Player A, Player B)	Cooperation (C)	Defection (D)

Cooperation (C)	(R, R)	(S, T)
Defection (D)	(T, S)	(P, P)

In this experiment, the parameters selected were consistent with those that Axelrod [14] used in solving the prisoners' dilemma. That is to say, the incomes were $R = 3$, $T = 5$, $S = 0$, and $P = 1$, which satisfied the conditions: $T > R > P > S$ and $2R > S + T$.

In addition, each two strategy model was competed for N times, namely, both sides had to make N selections. The result of each selection was recorded and the selection information from one side was passed on to the other side. After N competitions, the results of the each two strategy model were listed in forms of their total score. According to the total score, the strategy of corresponding strategy models was evaluated. In general competition, N was set as 10,000. It can be seen that the game, in the long-term, could yield stable income results.

2.2. Typical Strategy Models

(1) TFT strategy, that is, "return like-for-like" strategy is the most famous model for prisoners' dilemma. The main idea of this strategy is that: by starting with cooperation, the strategy selection of a round is made on the basis of the selection from the previous round. That is, if the rival selects cooperation or defection in the previous round, the selection will be repeated in the current round. This strategy performed best in the artificial intelligence algorithm competition organized by Axelord [14] (although the TFT strategy in this study was consistent in concept with that TFT strategy, the difference here lay in its details).

(2) PTFT: an improved TFT strategy [15]. This strategy is more selfish than TFT. It still starts with cooperation, however, in the following rounds, cooperation is only selectable in the case of an absence of defection for three rounds.

(3) GTFT: another improved TFT strategy [8]. Its strategy allows a certain probability of cooperation in the case of rival defection, and a certain probability of defection in the case of cooperation. It solves the deadlock arising from mutual defection in the competition.

(4) Pavlov: a different strategic concept [8]. It bullys the weak and fears the strong, namely, cooperation is continued in cases of mutual cooperation. However, defection is selected when one side chooses defection. Moreover, in the case of mutual defection, cooperation is given priority. Such a strategy represents a local optimum in genetic algorithm terms.

(5) Random: random strategy, which is, randomly returning to cooperation or non-cooperation. In related programs, a 50:50 random strategy is more commonly applied, that is, the probability of returning to cooperation or defection is 50 %. This strategy is mainly adopted to assess fixed strategies, set competition parameters, *etc*.

(6) Normal: a strategy mode developed by simulating common players. In this strategy, cooperation or defection would be selected with different probabilities based on the selections of both sides in the previous game. This strategy simulates player participation in a game using different strategies.

2.3. The Bayes Model

This proposed model was essentially a strategy constructed using statistical machine learning principals. In detail, all game data of both sides before the current round were used as a training data set for learning to obtain the most probable attitude that would be selected by the rival and the player themselves. In the learning process, naive Bayesian classification was used throughout.

In a general strategy model, the current attitude of the players is mainly determined by the selections of both sides over previous steps. Therefore, the game attitudes of both sides over these steps can be used as the characteristics for classification. Here, for simplification, only the last one step was considered. A 2-dimensional characteristic vector $x=(x_1,x_2)$ is introduced, where, x_1 and x_2 belong to {C, D} and represent the selections (with probabilities of selecting cooperation (C) or defection (D)) of players A and B in the previous round respectively. Moreover, an output remark y (belonging to {C, D}) was used to represent selections that could be cooperation (C) or defection (D). For each game, a data set (x_i,y_i) can be yielded by recording the selections of both sides in the previous round (x_i) and the selection of rival in the current round (y_i). As a consequence, before the N^{th} game, the training data set $T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$ was obtainable. According to the aforementioned data set and Bayes' formula [16]:

$$P(Y=c_k|X=x)=\frac{(P(Y=c_k)\prod_j P(X^j=x^j|Y=c_k))+1}{(\sum_k(P(Y=c_k)\prod_j P(X^j=x^j|Y=c_k)))+|c_k|}$$

Where, c_k refers to the situation that may appear in the previous game, $c_k=4$. The formula above gives rise to a cooperation probability sequence $\{P(Y=C|X=x)\}$, which denoted the current statistical data. The probability of the rival selecting cooperation was calculated in the case of different strategy selections of both sides in the previous step.

Subsequently, a k^{th} strategy selection chain was established to simulate the strategic selections in the following k steps to acquire the optimum solution. The algorithm steps were: firstly, all k^{th} step strategies that can be selected by player A $s_X=\{x_1,x_2,\cdots,x_k\}$ were generated. For each strategy of player A in the k^{th} step, player B will produce several such k^{th} step corresponding strategy chains $s_Y=\{y_1,y_2,\cdots,y_k\}$. However, since the strategy selected by player B was affected by the selection of player A in the previous step, the strategy chains had differing probabilities of selection. Moreover, considering that different strategies result in different incomes [1], the income expectation of each k^{th} step strategy chain of player A should be solved using the formula below:

$$EX=\sum val_X(S_X,S_Y)*P(S_Y=s_Y|S_X=s_X)$$

Where, *X* and *Y* refer to the strategy chains of the selections made by players A and B respectively; $val_X(S_X,S_Y)$ denotes the income of player X when the strategy chains of players A and B were S_X and S_Y respectively; $P(S_Y=s_Y|S_X=s_X)$ represents the probability of the strategy chain of player B being *Y* in the case of the strategy chain of player A being *X*. Moreover:

$$P(S_Y=s_Y|S_X=s_X)=\prod_{i=1}^{k}P(Y_i=y_i|Y_{i-1}=y_{i-1},X_{i-1}=x_{i-1})$$

The equation above was based on the assumption that the decision of player B was only related to the previous step.

The strategy finally selected was that with the maximum expectation value in the k^{th} step strategy chain obtained using the method above that is:

$$X_{pick}=\arg\max_{X_K} EX$$

In the real model, the previous 100 games (1% of the total number run) adopted a random strategy to collect the strategic model data from player B. In the following games, the strategy above was used to make decisions.

2.4. The Multi-player Prisoners' Dilemma Model

In reality, people not only face a double-player prisoners' dilemma but multi-player variants thereon. Therefore, multi-player formats afford the opportunity for a useful expansion of the prisoners' dilemma game. By referring to a published multi-player dilemma study [17], a reward and punishment rule was defined for this study:

(1) When all players selected cooperation (C), their income R was averaged across each player.

(2) When partial players selected defection (D), their income T was averaged out among the treacherous players, while income S was shared amongst cooperatives players.

(3) When all players selected defection (D), the income P was averaged out among all players.

For a prisoners' dilemma with $n = 4$, the parameters were set to: $R = 12$, $T = 10$, $S = 0$, and $P = 4$, which was in agreement with the standard form of the game. That is, the individual optimum solution of every player was obtained when one player selected defection, while the overall optimum solution was obtained when all players selected cooperation.

In this prisoners' game, the four strategies of each group were unavailable to the other players before they made their decision. After decisions were made, the income of each player was calculated and the decisions were revealed: the game was repeated 10,000 times.

2.5. Strategy in the Multi-player Prisoners' Dilemma based on Statistical Machine Learning

Using the strategy model based on naive Bayesian classification, the random variable distribution sequences of the decisions of each player made in any current situation were attainable. This sequence was in fact a joint distribution containing n variables. Moreover, it can be assumed that these variables were mutually independent. Therefore, in the following strategy selection, the optimal strategy selection for the next round can be obtained by merely using Bayes' theorem and analyzing the distribution sequence of the selections made by other players in the current situation.

2.6. Evaluation of the Strategy Model

In this study, each strategy model was provided with an evaluation mechanism to judge their performances. The evaluation mechanisms were based on two assumptions:

(1) Strategy models were forgetful: the decision of each strategy was only affected by a few of the most recent games. In the present study, the memory of each strategy was set to one.

(2) Strategy models were selfish: the relative income of an individual player was the prime driver for each strategy model.

These two assumptions facilitated the establishment of an evaluation mechanism of each strategy model and it followed the law of partial problems in real life. Using the evaluation mechanism obtained the strategy models can be more effectively assessed through a large number of repeated experiments. Simultaneously, according to the total income of each strategy model obtained in multiple runs of the game, the influences of each strategy model on the overall game results could be more objectively evaluated.

During the experiment, decision uncertainties were inculcated to build more realistic models. In other words, each strategy model was provided with certain probabilities by which they could randomly select either cooperation or defection. Meanwhile, such modification reduced the probability of a deadlocked decision (*i.e.* once two TFT models were mutually cooperative, the cooperation will last forever). In this experiment, each strategy was assigned a probability of 5 % with regards such decision-making.

3. Experimental Results and Analysis

3.1. The Performance of the Double-player Strategy Model: Naive Bayesian Classification

Figure 1 shows the overall incomes of both players recorded over 10,000 games comparing the proposed Bayes model with the other four typical models. Overall the Bayes model was more advantageous and achieved a higher score (overall income) than the other four. Of these other four typical strategy models, TFT performed best. It showed an equivalent overall income comparable to that of the Bayes model and a higher income than all the others. For each game pair, two stable test results, corresponding to the game score of the selected game pair, were presented here where they were just treated as two sets of solutions to this game. Further analysis suggested (Figure 2) that the Bayes model earned a higher income than the random, Pavlov, and GTFT models. The income ratio to the GTFT model reached 6.6, while that with the TFT model also exceeded one.

Subsequently, the occurrence times of each game and that of the Bayes model scoring 5, 3, 1, and 0 were statistically analyzed. As shown in Figure 3, the scores of 5 and 1 represented a relatively large proportion. That is to say, the Bayes model was prone to defection. Analysis of Figures 1 and 3 implied that the high scores were mostly close to five. Moreover, the results of each game competition showed that the Bayes model gained more when manifesting its tendency to defection.

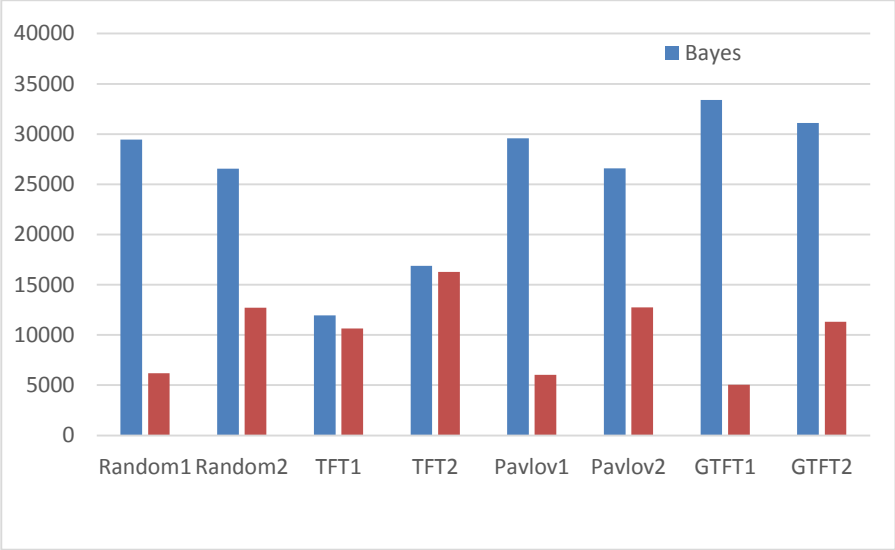

Figure 1. The Income of the Models after 10,000 Games

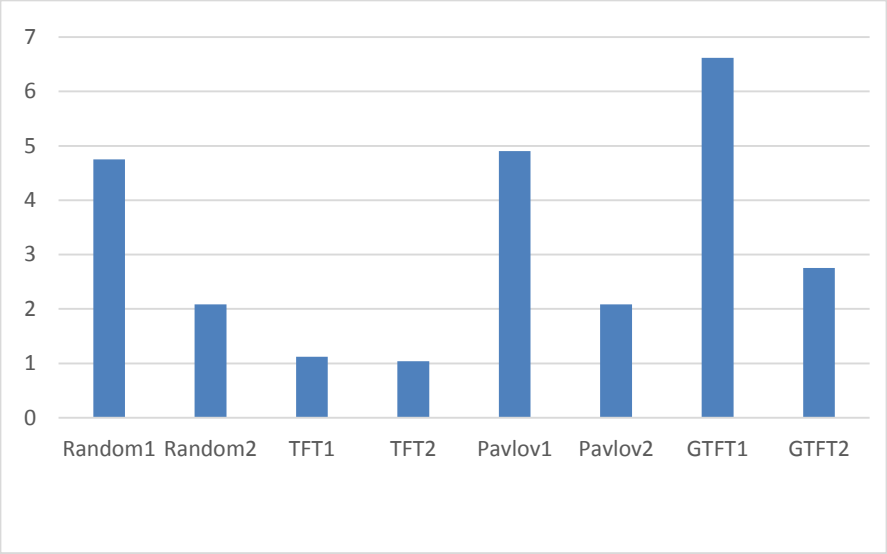

Figure 2. The Income Ratio of the Bayes Model to other Models

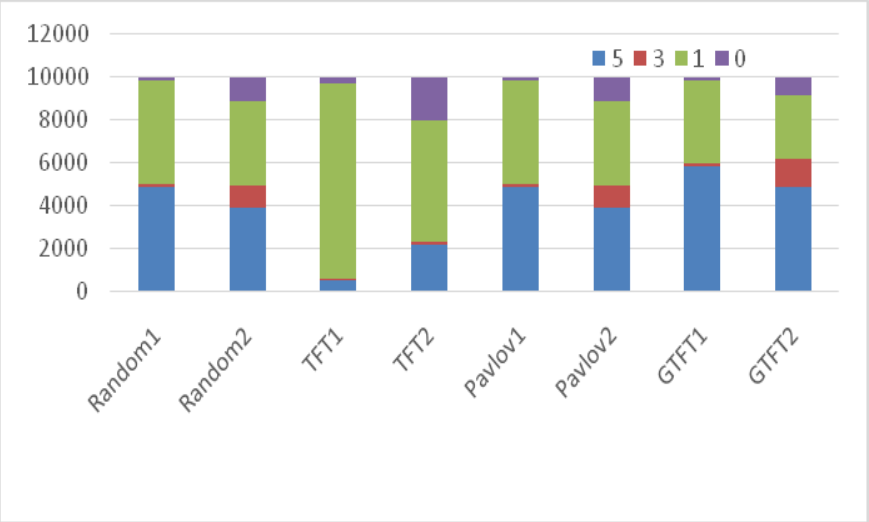

Figure 3. The Distribution of Bayesian Scores in Each Game

Analysis of the overall income of both players in each game (Figure 4) showed that the overall income in the game with a TFT model was lower. According to the performance of the rivals in each game (Figure 3) and comparison with test result 2, it was noted that both strategy models in test result 1 were less inclined to defection. Therefore, their overall income was higher. Since TFT is considered to be the model that can achieve more desired results amongst the four typical strategy models assessed here, games setting the TFT strategy model and the Bayes model in opposition were mainly investigated. In this game, the overall incomes of both sides were lower than those in other game competitions. This indicated that the game between the TFT strategy model and the Bayes model suffered simultaneous losses. By studying the single game results from the Bayes model and the TFT model, it was found that the scores from the Bayes model were mostly close to one (both sides simultaneously selected defection). This result suggested that in any game between the Bayes model and the

TFT model, defection appeared more frequently and represented a mark of the defection-prone tendencies of the two models.

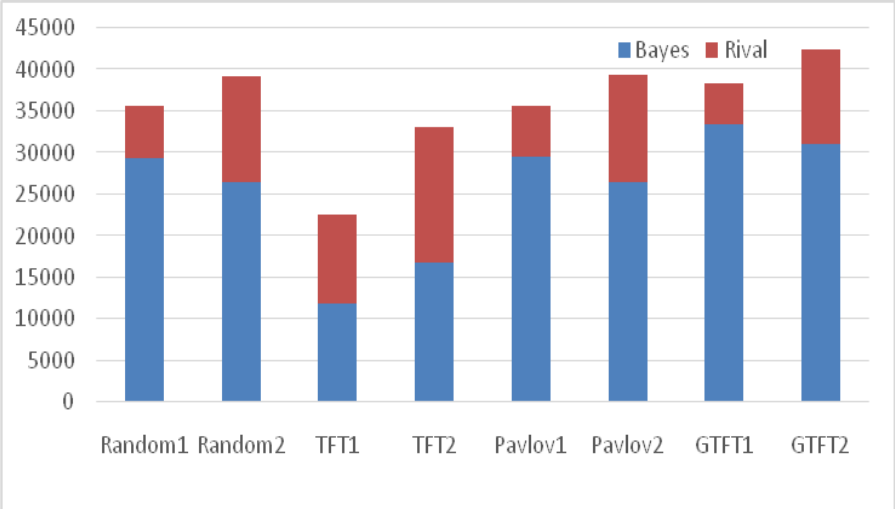

Figure 4. The Overall Game Incomes

3.2. The Performance of the Multi-player Bayes Model

With regard to multi-player games, this study used four different strategy models simultaneously. The overall income from each model was recorded, the income accruing to each player was distributed according to the protocols described in Section 2.4, and the overall income was calculated. In this section, the decision method of TFT, Pavlov, and GTFT models differed slightly from those applying to the two-player game: in the event of the defection of one of the other players in the previous round, the rivals selected defection. In the following two-player game, the decision for the current round was made by investigating the decisions of players A and B in the previous round.

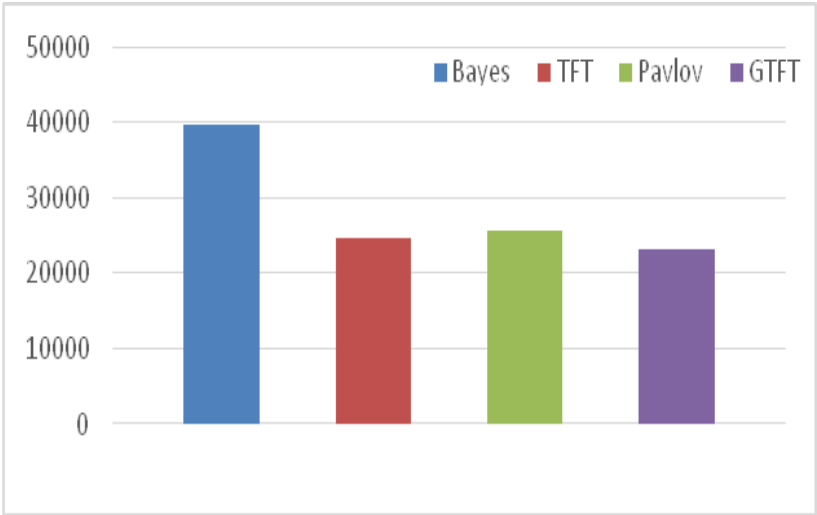

Figure 5. The Overall Income of Each Model after 10,000 Times of Multi-player Game

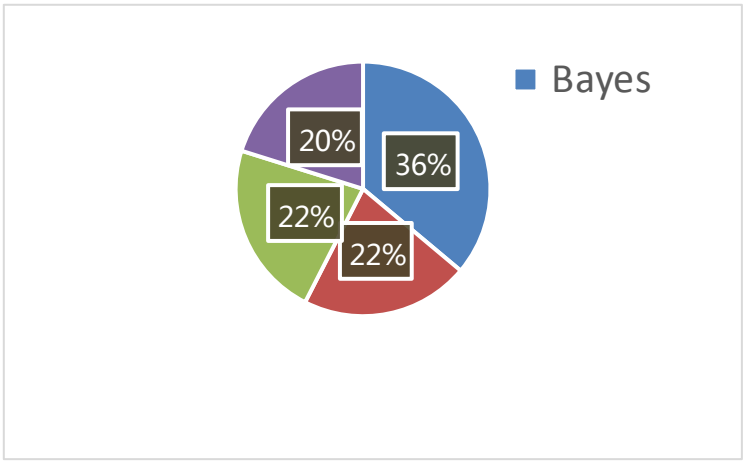

Figure 6. The Proportion of the Income of Each Model in the Overall Income of all Players

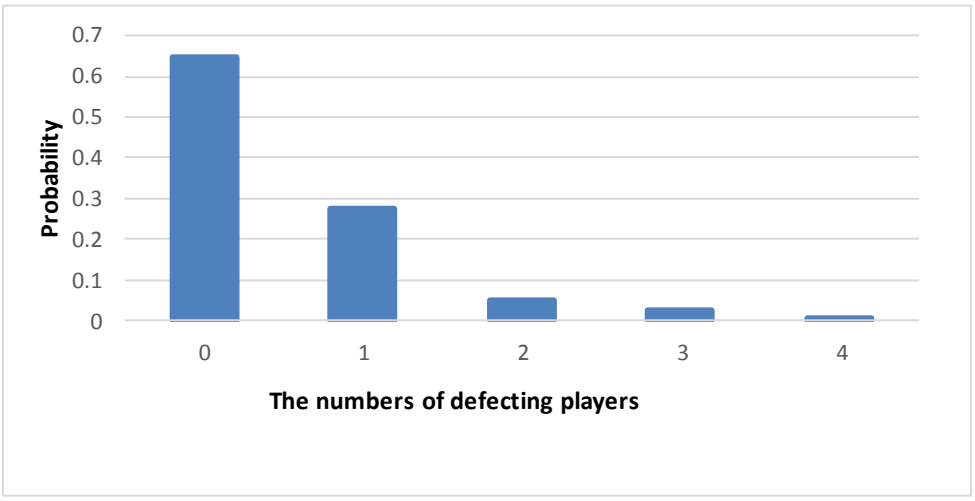

Figure 7. The Probability of Different Numbers of Defecting Players

It can be deduced from Figures 5 to 7 that, in the multi-player game, the Bayes model returned the highest overall income. The overall income of the Bayes model accounted for 36 % of the overall incomes of the four models. In addition, the overall game situation implied that the proportion of times that the four typical strategy models selected cooperation was the highest. That is to say, in the game with four models, each treated cooperation as its main strategy.

3.3. Analysis of the Performance of the Bayes Model versus Common Models

The general model refers to the strategy models that are possibly encountered in real-life enactments of the game, simulated here using a natural model. The natural model was a model adopting a random strategy (that is, when faced with identical decisions from the previous round, the probabilities that the natural models selected cooperation were different). To verify that the strategy selected by the proposed Bayes model reaped more benefits in games *versus* the general model, the competition between them was repeated 500 times. In

each game, there were 1,000 selections. It was an attempt to more comprehensively analyze the advantages and disadvantages of the Bayes model.

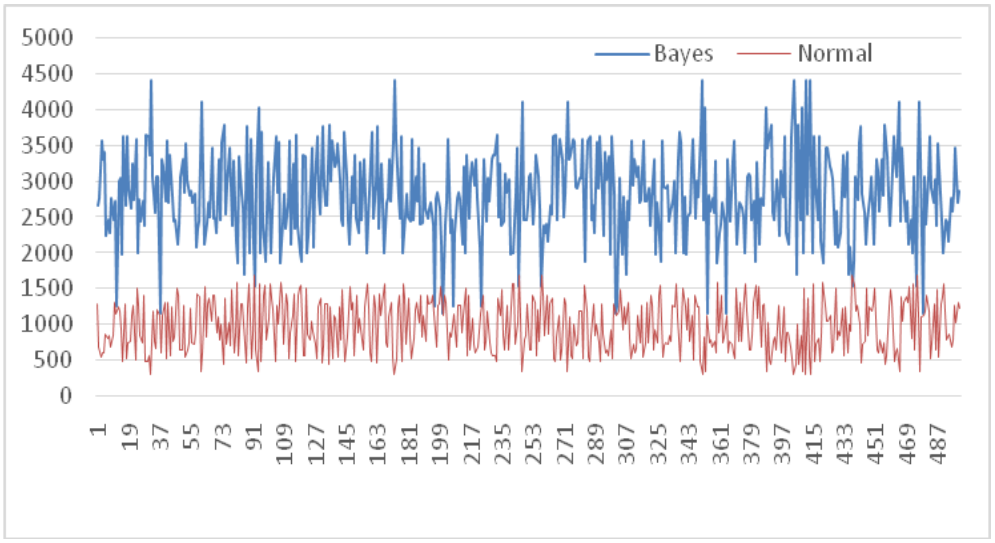

Figure 8. The Income of the Bayes Model and General Models in the Game

Figure 8 shows that the Bayes model performed better than most of natural models. Overall, the income from the Bayes model can reach approximately 3,000 and even 4,500 in individual extreme cases. The income ratio in Figure 9 was maximized at approximately 14, while most of the income ratios were above one.

In addition, the TFT model was also put in competition with the natural model: the result is shown in Figure 10. It was shown that the average income of the TFT model was approximately 2,500, which was lower than that of the Bayes model, while still equivalent to those of its other rivals.

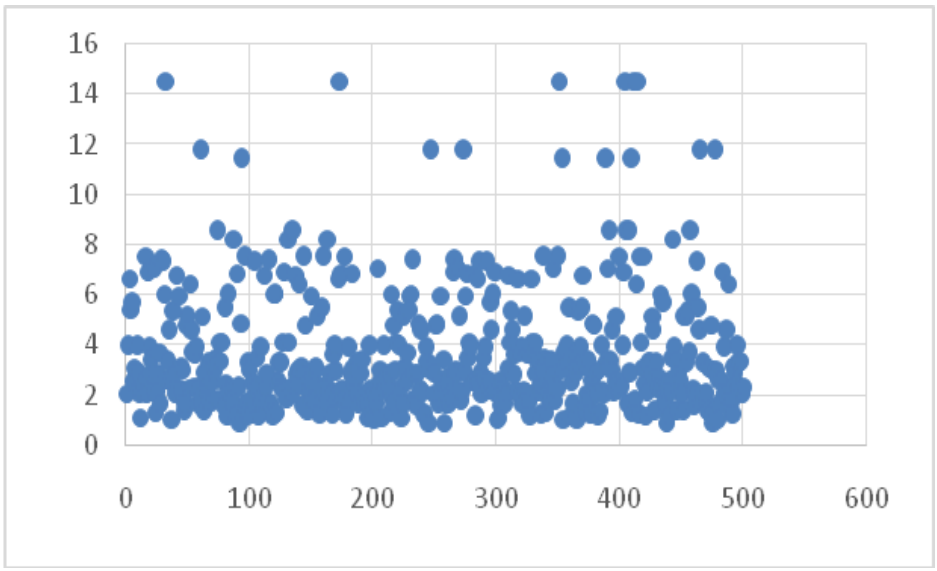

Figure 9. The Income Ratio of the Bayes Model *versus* the General Model

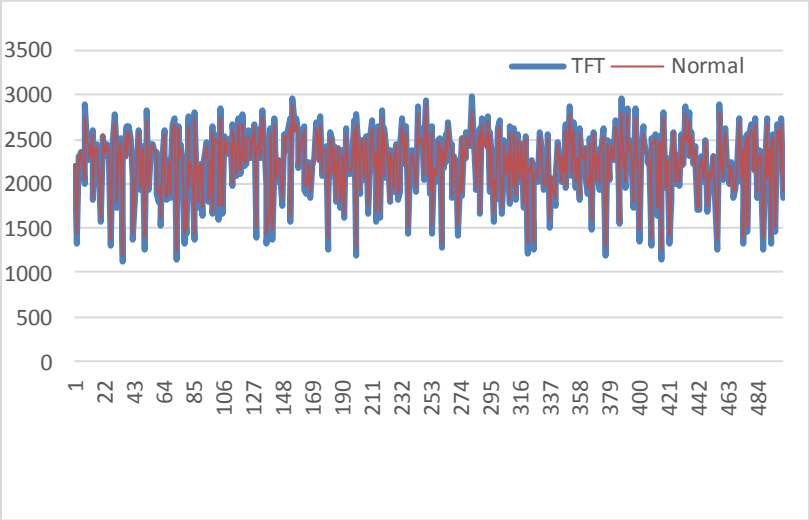

Figure 10. The Income of the TFT Model and General Models in the Game

3.4. The Performance of the Bayes Model when Run over Fewer Games

Since the Bayes model was a machine learning model, it needed a certain amount of data to guarantee its learning. Therefore, under fewer game runs, whether or not the Bayes model could achieve better game results was given careful consideration.

In this study, the Bayes model was evaluated using fewer game times in competition with the TFT model 100 times with the game repeated 100 times. Figures 11 and 12 show that the Bayes model was more advantageous.

An experiment was also carried out using a mere 10 games between the Bayes model and its TFT opponent. Figure 14 shows the fixed points obtained from this experiment. The appearance ratio of each point can be found with the help of Figure 13. As shown, a draw (the condition whereby both models got equal income) was the most common result, followed by a win for the Bayes model (*i.e*. higher income), and then the probability of higher income for TFT model was least.

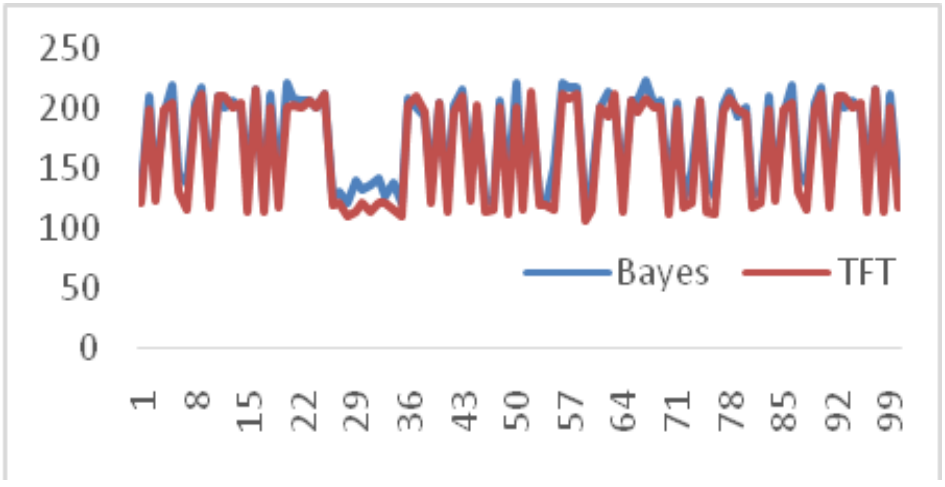

Figure 11. The Incomes of the Bayes and TFT Models after 100 Games

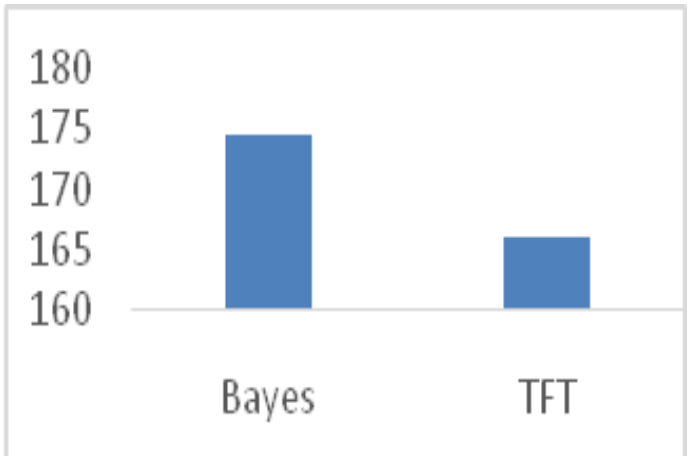

Figure 12. The Average Income of the Bayes and TFT Models after 100 Games

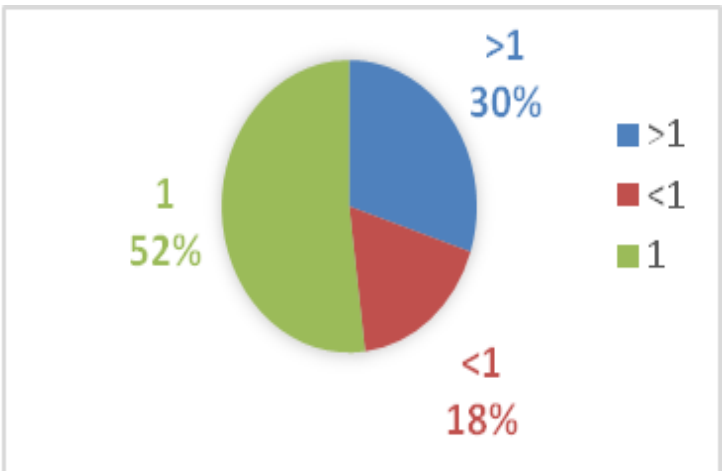

Figure 13. The Income Proportion Distribution of the Bayes and TFT Models after 10 Games

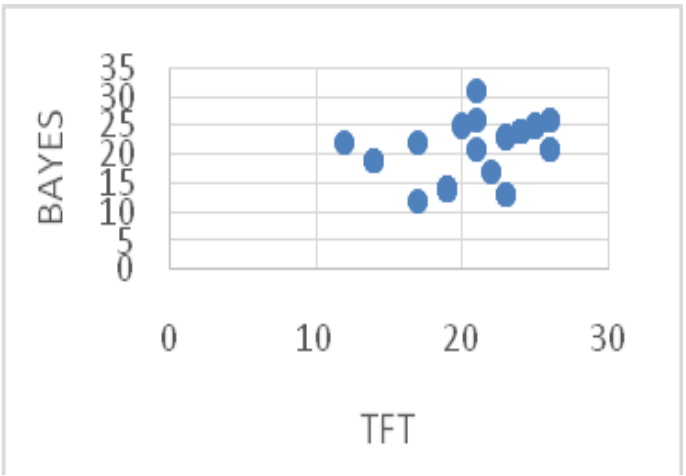

Figure 14. The Income Distribution of the Bayes and TFT Models after 10 Games

3.5. The Game Results from a PTFT Model Compared with the Other Models

The game models discussed above merely considered the attitude during and immediately after the selection of the previous step of both sides, while the PTFT model took account of the attitude during the selection of the previous three steps. Over 10,000 runs of the PTFT model against the other models, the incomes of each model are shown in Figure 15 which suggested that the Bayes model was disadvantageous over the game and gained neither more nor less than the TFT model (both models became trapped in the mutual defection deadlock). Moreover, the Pavlov model gained the least, while the GTFT model gained the most.

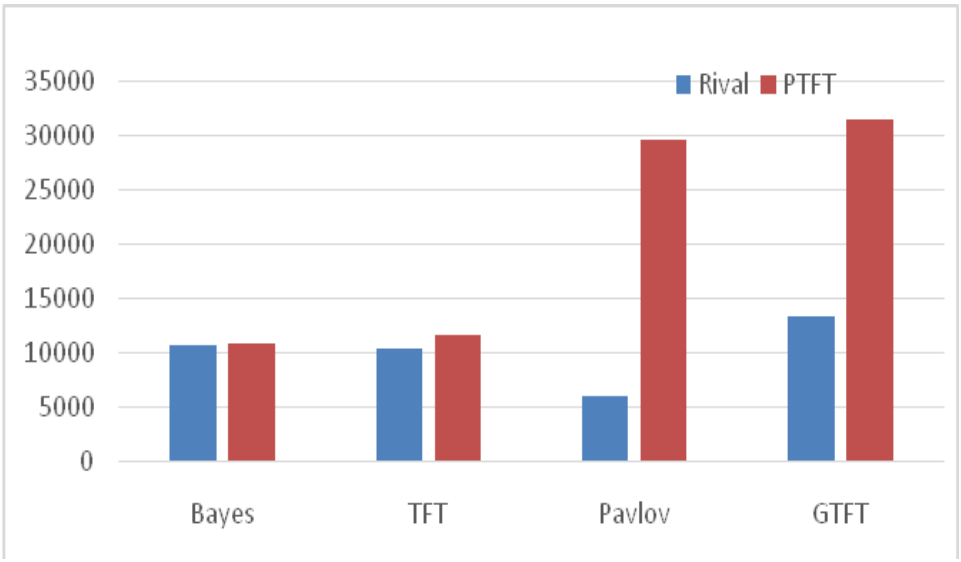

Figure 15. The Incomes of Each Models after 10,000 Games

This revealed one of the disadvantages of the Bayes model: it failed to comprehensively consider all state characteristics that may appear in the game. Therefore, the optimum solution was unattainable. In the current experiment, since the decision state steps of both sides considered by the Bayes model were set to one, the Bayes model was incapable of obtaining the optimum income result in the game against the PTFT model.

4. Conclusions

The trials suggested that the Bayes model proposed in the present study can yield more desired game results compared with its conventional counterparts. It was even believed that the Bayes model was slightly better than the acknowledged optimal strategy TFT model. In the game with more general single-step decision modeling and fewer games runs, the Bayes model also dominated. This result indicated that the naive Bayesian classification algorithm was feasible and effective at establishing the strategy model of an incomplete information game. It provided a novel idea for solving such a game.

However, the results obtained by the naive Bayesian classification algorithm showed certain defects: it was unable to obtain the desired solution in the case of the decision ability of a rival beyond its estimation range. Therefore, it reduced the applicability of

the machine learning algorithm when encountering complex models. This is should be the subject of future research.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (No. 61100148), and the Research Project on the Integration of Industry, Education and Guangdong Province (No. 2012B091100489).

References

[1] Nomia, "Games with Incomplete Information", UniversitÃ© PanthÃ© on-Sorbonne, **(1998)**.
[2] Harsanyi, "Games with Incomplete Information Played by "Bayesian" Players", The Basic Model&. Management science, **(2004)**.
[3] M. M. Zinkevich and M. Bowling, "Adv. in Neur. Inf. Proc. Sys.", vol. 20, **(2008),** pp. 1729-1736.
[4] E. Alpaydin, "Introduction to machine learning", The MIT Press, **(2004)**.
[5] M. CBishop, "Pattern recognition and machine learning", Springer, New York, **(2006)**.
[6] X. Zhang, "Acta Auto. Sini.", vol. 1, **(2000),** pp. 32-42**.**
[7] R. J. Hanson and C. P. Stutz, "Bayesian classification theory, Artificial Intelligence Research Branch", NASA Ames Research Center, **(1991)**.
[8] M. Nowak and K. Sigmund, "Nature", vol. 6432, **(1993),** pp. 56-58.
[9] D. MP Kreps and J. Roberts, "J. of Eco. Theo.", vol. 2, **(1982),** pp. 245-252.
[10] D. W. K. Yeung, L. A. Petrosyan and M. C. C. Lee, "Dynamic Cooperation: A Paradigm on the Cutting-edge of Game Theory", China Market Press, **(2007)**.
[11] Axelrod, "J. of Conf. Reso.", vol. 1, **(1980),** pp. 3-25.
[12] Miller, "J. of Econ. Beha. & Orga.", vol. 1, **(1996),** pp. 87-112.
[13] H. Lin and C. Wu, « Acta Physica Sinica ", vol. 8, **(2007),** pp. 4313-4318.
[14] A. R W D. "Hamilton, Science", vol. 4489, **(1981),** pp. 1390-1396.
[15] S. R. R. Stoecker, "J. of Econ. Beha. & Orga.", vol. 1, **(1986),** pp. 47-70.
[16] F. P AN Lachiche, "Mach. Learn.", vol. 3, **(2004),** pp. 233-269.
[17] D. Jiang, "Situation analysis of double action games with entropy", Science Press USA Inc., New York, **(2010)**.

