THEORETICAL FRAMEWORK TO LEARN OPPONENT'S MIXED STRATEGY: PLAYING "PAPER, SCISSOR, ROCK" BY EVOLVING ADAPTIVE NEURAL NETWORKS

M. S. Khan, W. M. Qazi, S. Lodhi

Physics Department, GC University, Lahore.

ABSTRACT: This paper reports a theoretical framework that enable an artificial player to observe game played between two human players, then to learn the hidden patterns of pure strategies in their selected mix strategies. Finally, when artificial player (agent) plays with a particular human player, to whom it has already observed playing, artificial player takes its first move by selecting encounter strategy for its opponent's dominating strategy. Framework uses genetic algorithms to evolve adaptive neural networks. Framework will help in designing artificial players (multi agents) capable to meet basic assumptions of game theory i.e. common knowledge of rationality and consistent alignment of beliefs.

Keywords: *Artificial Neural Networks, Game Theory, Genetic Algorithms, Intelligent Agents, Machine Learning.*

INTRODUCTION

Game theory oriented systems generally have tendency to attain equilibrium. This behavior results when player makes self assessment about the situation. Self assessment is made when the rules of the games, rationality of the players and player's payoff function are prior and common knowledge. Indeed there are certain cases where such tendency is not expressed (Fudenberg and Levine, 1998; Aumann, 1992; Htun, 2005). Furthermore, such behaviors (shift from non-equilibrium to equilibrium) do not align with the introspective theories (Fudenberg and Levine, 1998).

Another approach to resolve the scenario is that players must learn strategies (Sato et al, 2002). One of the significant characteristic of learning dynamics is that it provides means to attain equilibrium (Shapley, 1964) and prevents the game from getting chaotic (Cowen, 1992). Paper-scissor-rock (PSR) game is a common instance of such systems. This game requires skills to play if the game extends over many episodes. This recurrence of episodes provides a mean to predict and exploit the behavior of an opponent (http://en.wikipedia.org/wiki/Rock%2C_Paper%2C_Scissors). Certainly under learning systems there are chances that PSR game may fail to converge to equilibrium (Sato et al, 2002)

World Rock–Paper–Scissors Society web site (www.worldrps.com/gbasics.html) suggests that members of this society do not play the Nash equilibrium strategy. They use psychology to handle the opponents move by inducing responses in the opponent. This means that real people instead of acting rational they try to learn the opponent's strategy.

Real challenge arises when one of the opponents is not human. The challenge is to develop artificial player which can play against human player (like in traditional computer games). In recent years convergence between game theory and artificial intelligence has been observed. Few of the important characteristic of artificially intelligent system is that it can learn, adapt and evolve.

Number of methods has been reported in past which deals with these characteristics, e.g. artificial neural networks, genetic algorithms, Bayesian networks, etc (see Russell and Norvig, 2004). Evolutionary neural networks (ENN) are one of these promising methods. ENN provides a mean to evolve number of neural network individual and then selects fittest among them (Zhang and Muhlenbein, 1993; Braun and Weisbrod, 1993; Braun el, al., 1996). Different variations of ENN have also been reported with the feature of adaptive learning. In such networks, learning rate is adaptive. Learning rate in adaptive strategy is larger in the beginning of the training session. As training proceeds, learning rate gets smaller, if total error is reducing. Number of strategies has been reported to implement adaptive learning rate (Janakiraman, 1993, Vogl. et at, 1988, Mackay, 1991, Niklasson et al., 2001). Moreover strategies for selecting sample size and to control evolution complexity in genetic algorithms has also been reported (Gao, Y., 2003).

Decision making in games like PSR are based on the models learned by the player by observing series of temporal events. With reference to this fact, the task is to predict opponent's next move to encounter the opponents move. Such prediction needs to take into account the previous moves of all players of the game. If an artificial player is needed to be designed based on artificial neural network which can play PSR with human player then recurrent and tapped line delay neural networks (TDNN) are possible candidates. Basic structure of TDNN is shown in figure 1. Work has been reported previously which haves shown the significance of the adaptive tapped line neural (Niklasson *et al.*, 2001)

This study reports a theoretical learning framework as an extension of the concept reported previously (Niklasson *et al.*, 2001). This framework "PSR Evolutionary and Adaptive Learning (PSR-EAL) Framework" provides a standard to design artificial player that can play PSR with its human opponent(s) by learning their strategies, combined with its experience of playing with humans using evolutionary and adaptive neural networks.

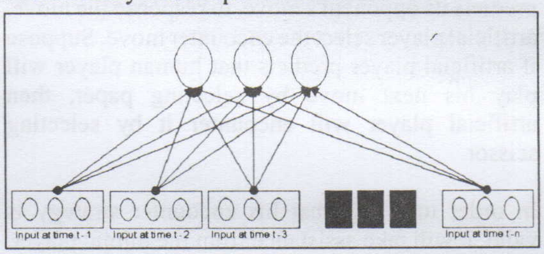

Fig 1: Simple Tapped Line Delay Neural Network

PSR-EAL FRAMEWORK

PSR-EAL Framework provides artificial player (AR) a mechanism to choose next move, based on past moves (for example 5 moves) played in current game, experience with opponent's game playing psychology (i.e. how opponent thinks) and the way in general the way humans choose their next move. Each of these mentioned tasks are modeled in various functional sub systems of PSR-EAL framework. These individual functional sub systems try to achieve their goals by under going into evolution of their own internal information states, structure and configuration through learning by observation and experience using "Evolutionary Adaptive Tapped Line Neural Networks".

A game under this framework is defined a collection of "n" number rounds played by two particular players. A game can be played between human player vs human player / human player vs artificial player. Any player p is said to be a winner if the number of rounds won by it are greater than the number of rounds won by the opponent. Two opponents can play multiple games with each other consecutively or at different times. When game is played between two human players, then artificial player sits and observe the game. The objective of this observation is to learn; the way both human players think and take decision, to select their next move. This task is achieved by playing mimic games in observatory environment.
When game is played between human player and artificial player, the artificial play switches it self to play battle game in combat environment. A Game and its environment is loaded and controlled through game server. All participants get them self connected to game server using their player identity. A potential player may get his/her player identity through signup process.

Mimic Games and Observatory Environment
Observatory environment: Observatory environment is a gaming environment. In this environment two human player play game against each other using their remote terminal machines. These remote terminals are connected with game server as shown in figure 2. Artificial player is also connected with the game server as game observer. As soon as two human players decide to play game with each other, they load observatory environment from game server. When observatory environment is loaded, artificial player switches it self to mimic game playing mode. Every event occurring during the game is broadcasted as by game server. This broadcasted information is treated as acknowledgment at human playing terminal end and observatory information for artificial player end.

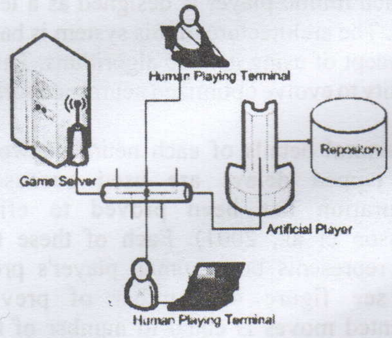

Fig 2: Observatory environment and mimic game

Mimic Games: In mimic games, artificial player perceives the observatory information, broadcasted by the game server. This observatory information in processed by artificial player's mimic engine, resulting into learning updates for each human player.

Mimic engine is consisted of two mimic players corresponding to each human player as shown in figure 3. Mimic player (A) impersonates human player A and similarly, Mimic player (B) impersonates human player (B). The objective of each mimic player is to learn the decision making behavioral patterns, corresponding to its human player, using observatory information. Consider at any stage during the observed game, mimic player forecasts that its corresponding player will, for example, choose paper in next move; and the human player does select paper in his next move; then it will be said that mimic player is now imitating its human player and has accomplished its learning objective. This imitating will help artificial player to think in the same way as its corresponding human player may be thinking.

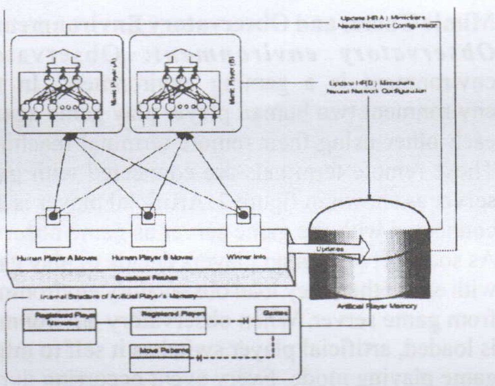

Fig 3: Architecture of mimic games with mimic player encapsulated in artificial player.

Mimic Player Design: To achieve this complex task, each mimic player is designed as a learning system. The architecture of this system is based on the concept of using genetic algorithms. This give the ability to evolve optimized neural network.

Architectural details of each neural network are series tapped delays are used because this configuration has been proved to effective (Niklasson et al., 2001). Each of these tapped delays represents both human player's previous move see figure 4. Number of previously represented moves is equal to number of tapped delay neurons in the configuration of the neural network. In each time step artificial player predicts the next move of its corresponding human player. This move is represented by the unit with higher activation value (often called winner-takes-all strategy) because the threshold function is linear

BATTLE GAMES AND COMBAT ENVIRONMENT

In battle games, artificial player plays with human player as his/her opponent. During battle environment games places both player in combat environment. Before playing with human, artificial player first checks that had it observed this human player playing game before in observation environment. In other words, artificial player searches its database to see whether it knows how this human player plays game. If artificial player has already observed playing game in past then, it loads the opponent human player(s) mimic player as shown in figure 4.

In first move of the game, artificial player takes its dominating move by selecting at random (paper, scissor or rock). In second play using the previous moves, artificial player will predict next move using adaptive neural network. Prediction mechanism is same as explained in previous section; expect during battle game there is only one neural network. In battle mode neural network predicts its opponent's move, to response the move, artificial player select the encounter move. Suppose if artificial player predicts that human player will play his next move by selecting paper, then artificial player will encounter it by selecting scissor.

In order to verify that his encounter strategy is valid, it will take assistance from the mimic player. For each move mimic player will mimic human player. If the result from mimic player and the predicted once are same then, artificial player will simply play the predicted move's encounter.
If there is difference between artificial players has predicted and mimic player, then artificial player will compute beliefs to choose which move to play in the conflict. Beliefs are calculated using Bayesian probability (Heap and Varoufakis, 1995; Russell and Norvig, 2004).

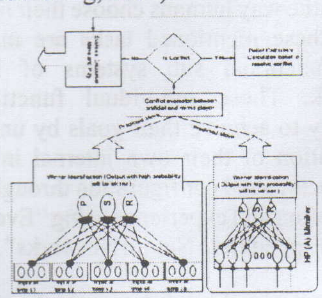

Figure 4: Architecture of artificial player in combat games

Bayesian probability will be computed for both moves, move with maximum Bayesian probability will be selected. Like this artificial player will play in combat mode.

DISCUSSION

PSR-Framework is designed keeping in view the aspects and assumption of game theory. Fundamental assumptions in game theory (Heap and Varoufakis, 1995) deals with following problems (i) How people selects an action, (ii) what we should assume about what motivates each player and (iii) most importantly, a tricky issue, what each thinks the other player will do in any set of circumstance. These issues gave birth to the basic assumptions of game theory i.e. Common knowledge and Consistency of Alignment in Beliefs. Key to play game is that one should be able to predict what his/her opponent's strategy in a given situations. In general these predictions are based on learned knowledge and experience. Learning performance in human is of increased by observing a task being performed (Bentivegna and Atkeson, 2001). It is suggested that machine which uses observation mechanism can enchance their learning usability and functionality (Bentivegna and Atkeson, 2001). For example, if machine could learn, to play a 'scissors, paper, stone' by observing and mimicking human players then this could lead to machines that can automatically learn how to spot an intruder or perform vital maintenance work (Knight, 2005). Number of studies has reported the framework to design robot that learn and play game by observation (Bentivegna and Atkeson, 2001). This work has reported a framework that enables, artificial player to learn they way his opponents take decision by observing his/her playing. As mentioned above, knowing how the opponent takes his decision is very important to win. Artificial player does this job using its sub system called human mimic player. When observing the human player, mimic player tries to learn their behavior and patterns in their decision making. This task is accomplished by learning using evolutionary adaptive neural networks. Mimic player evolves number of neural networks in its internal memory purpose of this evolution is to increase learning process. It has been reported that by evolving neural networks, learning process is increased and hence it converges (Vogl et al., 1988). When artificial player, in combat mode, plays game with the human player, whose pattern already it has already learnt, artificial player load its mimic player. In first move, artificial player selects random move in order to play it's dominate strategy and also to learn what dominating strategy does human player have to play. According to game theory assumptions; each player is rational and will always try to play initial game with his/her dominating strategy, because his/her objective is to minimize the risks involved in repeated games to move toward the rational strategies (Heap and Varoufakis, 1995). It is also reported that to play game like PSR best dominating strategy is to be random (Sato et al., 2002). Due to this reason framework provides enforces artificial player to take his/her first move on random basis. For next move, mimic player, taking into account the opponent's all previous games which he/she had played with other human, tries to predict his/her next move. Moreover, just keeping into account the current game's move, artificial player tries to predict opponent's next move. In other words, artificial player predicts two moves for opponent human player, one on the basis of current game and other on the basis of mimic player. These two predicted moves are then compared, if difference between these moves exist, then artificial player computes beliefs for both moves. Move with maximum belief is selected and played. This behavior of artificial player reflects that artificial player is capable to reason with incomplete information, noisy information and/or in uncertainty by computing it's belief about the opponents strategy. Game theory itself speaks about taking decisions and game playing when information is incomplete or uncertain (Heap and Varoufakis, 1995). The proposed model for computing beliefs is to use Bayesian rule (Heap and Varoufakis, 1995).

CONCLUSION

This work reports a theoretical framework using the best practices reported in previous studies. On the basis of the above discussion we suggest that artificial player, playing with PSR-Framework, will play rationally. When games are played with rational attitude and learning they converge and possibilities exist that they may attain Nash Equilibrium. Therefore, underlying framework motivates the artificial player to play rationally hence attain Nash Equilibrium at a certain time during game. In future we intend to implement and develop this framework. Framework in developed form will help in understanding some important concepts in game theory.

REFERENCES

Aumann, R. Irrationality in Game Theory, Economic Analysis of Markets and Games, Essays in Honor of Frank Hahn (1992).

Bentivegna, D. C. and Atkeson, C. G. Learning from observation Using Perimitives, Proceeding of IEEE International Conference on Robotics & Automation, Seoul, Korea. 1988 – 1993 (2001).

Braun, H. and J. Weisbrod, Evolving neural feedforward networks, in R.F. Albrecht (ed.), Proceedings of the International Conference on Artificial Neural Nets and Genetic Algorithms, Springer (1993).

Braun, H. and T. Ragg, Evolution of Neural Networks, User Manual and Implementation Guide, Technical Report 21/96, University of Karlsruhe (1996).

Cowen, S., *Doctoral Dissertation.* Berkeley: Univ. of California (1992).

Fudenberg D. and D. K. Levine The Theory of Learning Games, The MIT Press, Cambridge, Massachusetts, London (1998).

Gao, Y., Population size and sampling complexity in genetic algorithms, Proceedings of the Bird of a Feather Workshops (GECCO2003) ---Learning, Adaptation, and Approximation in Evolutionary Computation (2003).

Heap, S. P. H. and Y. Varoufakis, Game Theory: A Critical Introduction, 1st edition, Routledge, London and New York (1995).

Htun, Y. Irrationality in Game Theory, University of British Colombia, www.cs.ubc.ca/~kevinlb/teaching/cs532a%20-%202006/Projects/YaminHtun.pdf (2005).

Janakiraman, J. Adaptive Learning Rate for Increasing Learning Speed in Backpropagation Networks. Masters Project Report. Department of Computer Science. Iowa State University (1993).

Knight, W. Machine learn games like humans, New Scientist, New Service, January 24, 2005 (2005).

Mackay, D. J. C. Bayesian methods for adaptive model, PhD thesis, California Institute of Technology (1991).

Niklasson, L., H. Engstrom, and U. Johansson, An Adaptive 'Rock, Scissors and Paper' Player Based on a Tapped Delay Neural Network, Proceedings of the International Conference on Application and Development of Computer Games in the 21st Century, Hong Kong, 22-23 November 2001, 69-73 (2001).

Russell, S. and P. Norvig, Artificial Intelligence: A Modern Approach, 2nd Edition, Pearson Education (2004).

Sato, Y., E. Akiyama, and J. D. Farmer, Chaos in learning a simple two-person game, Proc Natl Acad Sci USA, 99(7), 4748-4751 (2002).

Shapley, L., Ann Math Studies, (5) 1–28 (1964).

Vogl, T.P., J.K. Mangis, A.K. Zigler, W.T. Zink, and D.L. Alkon, Accelerating the convergence of the backpropagation method. Biological Cybernetics. 59: 256-264 (1988).

Zhang, B., and H. Muhlenbein, Evolving Optimal Neural Networks Using Genetic Algorithms with Occam's Razor, Complex Systems, vol. 7 no. 3 pp. 199-220 (1993).