NEGOTIATION AGENT BEHAVIORS BASED ON REINFORCEMENT LEARNING APPROACHES AND FUZZY ARTMAP NEURAL NETWORKS

Amine Chohra, Arash Bahrammirzaee, and Kurosh Madani

Images, Signals, and Intelligent Systems Laboratory (LISSI / EA 3956), Paris-XII University, Senart Institute of Technology, Avenue Pierre Point, 77127 Lieusaint, France, {chohra, bahrammirzaee, or madani}@univ-paris12.fr, http://lissi.univ-paris12.fr

Abstract: *Behaviors, in which the characters conciliatory, neutral, or aggressive define a 'psychological' aspect of human personality, play an important role for negotiation agent. Elsewhere, learning in negotiation is fundamental for understanding human behaviors and developing new concepts. In this paper, a negotiation strategy essentially based on such human personality behaviors is suggested for SISINE project which aims to develop innovative teaching methodology of negotiation skills. For this purpose, first, reinforcement learning (Q-learning and Sarsa-Learning) approaches are developed, analyzed, and compared in order to acquire the strategy negotiation behaviors. Second, a Fuzzy ArtMap Neural Network (FAMNN) is developed to acquire this strategy. Third, a Field Programmable Gate Array (FPGA) architecture is suggested for the FAMNN integration. The suggested strategy displays the ability to provide agents, through a basic buying strategy, with a first intelligence level in a social and cognitive system for learning negotiation strategies (human-agent and agent-agent).*

Keywords: *Intelligent behaviors, decision-making, reinforcement learning, fuzzy artmap neural network, field programmable gate array.*

1. INTRODUCTION

In a context where agents must reach agreements on matters of mutual interest, *negotiation* techniques for reaching agreements are required. In general, any negotiation settings will have four different components [1]: - a negotiation set, the space of possible proposals that agents can make; - a protocol, the legal proposals that agents can make, as a function of prior negotiation history; - a collection of strategies, one for each agent, which determine what proposals agents will make; and - an agreement rule that determines the reach agreements stopping the negotiation.

Negotiation usually proceeds in a series of rounds, with every agent making a proposal at every round. The proposals that agents make are defined by their strategy (a mapping from state history to proposal; a way to use the protocol), must be drawn from the negotiation set, and must be legal, as defined by the protocol (which defines possible proposals at different rounds). If agreement is reached, as defined by the agreement rule, then negotiation terminates with the agreement deal. These four parameters lead to an extremely rich and complex environment for analysis.

Another source of complexity in negotiation is the number of agents involved in the process, and the way in which these agents interact [1]. First possibility is one-to-one negotiation, in which one agent negotiates with just one another agent, e.g., a particularly simple case where the agents involved have symmetric preferences with respect to the possible deals, e.g., when discussing terms with a car salesman. Second possibility is many-to-one negotiation. In this setting, a single agent negotiates with a number of other agents, and can often be treated as a number of concurrent one-to-one negotiations. Third possibility is many-to-many negotiation, where, many agents negotiate with many other agents simultaneously. In the worst case, where there are n agents involved in negotiation in total, making such negotiations hard to handle. For these reasons, most attempts to automate negotiation process have focused on rather simple settings. Single-issue, symmetric, one-to-one negotiation is the most commonly analyzed, and it is on such settings that we will mainly focus in this work.

In the research works developed aiming to analyze and describe the human behavior in [2], twelve categories representing three major parts of the behavior have been defined: the positive socio-emotional part, a neutral task part, and the negative socio-emotional part. In another side, in research

works on the social psychology of the negotiation of Rubin and Brown developed in [3], the interpersonal orientation of a person has an influence on his negotiating behavior. More, it is possible to manipulate the interpersonal orientation, at least indirectly, in various ways. For instance, the communication can be manipulated for example by varying the type and amount of interpersonal information made available to a party. This will influence the interpersonal orientation, which in turn will influence the quality of the negotiation. According to Rubin and Brown, the interpersonal orientation is predominantly concerned with the degree of a person's responsiveness. If he is not responsive, he stands to gain much in the negotiating situation due to the deliberateness of his behaviour. Responsive people are more co-operative and therefore expect positive results. The personality type should therefore be determined first to obtain the best results in negotiation. Elsewhere, a great number of successes in Artificial Intelligence can be attributed to a straightforward strategy [4]: linear evaluation of several simple features, trained by temporal-difference learning, and combined to an appropriate search algorithm. Games provide interesting case studies for this approach. In games as varied as Chess, Checkers, Othello, Backgammon and Scrabble, computers have exceeded human levels of performance.

Finding a shared solution to a problem within a group requires *negotiation*, a potentially exhausting and time-consuming process. To negotiate successfully, members have to involve the whole group, explain their position clearly and do their best to understand those of others [5], [6]. However, in reality, groups often fail to negotiate, even when negotiation would be useful for each part of the group. Sometimes the problem lies in sheer size of the group, or in hierarchical organizational structures or in impediments to communication deriving from language, culture or history. In other cases, the main barriers lie in the individual psychology of specific group members. Typical problems include weak communications skills, lack of empathy with others, and poor control over the emotions arising during prolonged discussion. Such problems can be directly or indirectly related to the personality of each group member participating to the negotiation. Thus, negotiation behaviors, in which the characters such as Conciliatory (Con), Neutral (Neu), or Aggressive (Agg) define a 'psychological' aspect of the personality of a negotiation member (negotiator), play an important role [2], [3].

Elsewhere, learning from interaction in negotiation is fundamental, from embodied cognitive science and understanding natural intelligence perspectives [7], [8], for understanding human behaviors and developing new solution concepts [9]. Learning from interaction is a foundational idea underlying nearly all theories of learning. Indeed, whether a human is learning to drive a car or to hold a conversation (during a negotiation), he is acutely aware of how his environment responds to what he does, and he seeks to influence what happens through his behavior. Elsewhere, reinforcement learning is much more focused on goal-directed learning from interaction than other approaches to machine learning [10], [11], [12]. More, reinforcement learning approaches offer two important advantages over classical dynamic programming [13]. First, they are on-line having capability to take into account dynamics nature of real environments. Second, they can employ function approximation techniques, e.g., Neural Networks (NN) [14], [15], [16], to represent their knowledge, and to generalize so that the learning time scales much better.

Humans have developed advanced skills in the intentions and the bodily expressions of the other human being, particularly important in high level communication which is at the basis of any ''successful'' negotiation (interaction process). Dealing with this, the SOcial and COgnitive SYStem, SOCOSYS, is developed for an intelligent human-agent interaction in our research laboratory. A part of this research work has been developed in [17] for SISINE[1] (Integrated System of Simulation for Negotiation) project and an extension of this work is developed in this paper.

In this paper, a brief description of SISINE project is given in Sect. 2. Then, a negotiation strategy essentially based on negotiation behaviors (characters Con, Neu, and Agg) of human personality is suggested for in Sect. 3 SISINE. For this purpose, first, reinforcement learning (Q-learning and Sarsa-Learning) approaches are developed, analyzed, and compared in Sect. 4 in order to acquire the strategy negotiation behaviors. Second, a Fuzzy ArtMap Neural Network (FAMNN) is developed to acquire this strategy. Third, a Field Programmable Gate Array (FPGA) architecture is suggested for the FAMNN integration in Sect. 5.

2. INTEGRATED SYSTEM OF SIMULATION FOR NEGOTIATION

In SISINE project, simulation environments enable a participant to interact with a virtual entity called ''bot'' (a software agent), through a communicative exchange: texts (one among three),

voice (tone and volume), facial expression, and gesture. The objective of such simulation is to allow participants to directly experience the basic elements of negotiation, through: - a ''standard bot'', an agent with reactions from simple rules, and a ''smart bot'' an agent with reactions from a first intelligence level. The suggested negotiation approach relies on a theoretical model of the negotiation process which is mainly based on the negotiator personality, i.e., characters Con, Neu, and Agg which define a 'psychological' aspect of the negotiator agent personality. Also, in this theoretical model, the character of a negotiator agent is defined by a character vector [Con, Neu, Agg] where each component belongs to the interval [0, 100] in percentage (%) such as: Con + Neu + Agg = 100 %. During a negotiation round, each negotiator agent is defined by its internal current state CurrentState(CurrentEscLevel, **CurrentChar**, **UserSentence**), having its CurrentEscLevel and **CurrentChar**, and receiving a sentence vector **UserSentence** = [DeltaEscLevel, CharToModify, DeltaChar] from a user where: - DeltaEscLevel, an escalation level variation from [-60, +60], escalation level defines gradually different negotiation stages from agreement to interruption with seven possible stages modeled by EscLevel from [0, 60] ; - CharToModify, a character to modify (Con, Neu, or Agg) ; DeltaChar, a character variation belonging to [-10, +10].

In fact, during a round (a given specific state) of a negotiation as illustrated in Fig. 1, an agent has a given CurrentEscLevel and a given **CurrentChar**, and receives DeltaEscLevel, CharToModify, and DeltaChar extracted in SISINE software from a user (human or another agent) sentence. An example of such negotiation session is illustrated in Fig. 1, where a conversation is shown, during a negotiation round, between two agents (woman and man) which are arguing about whether they should spend their holidays at seaside or in mountains. Goal for an agent, is from a user sentence to update its escalation level (NewEscLevel) and character (**NewChar**), and to choose an answer based on these new values.

3. NEGOTIATION STRATEGY

In this work, for an *intelligent* agent, given a state (round) and a given strategy (basic buying strategy), the agent reactions (choosing an answer in each state) must be based on the updating of the Qvalues (from reinforcement Q-Learning or Sarsa-Learning) corresponding to the negotiation behaviors. The basic three (03) negotiation behaviors (Con, Neu, Agg) have been extended to nine such as:

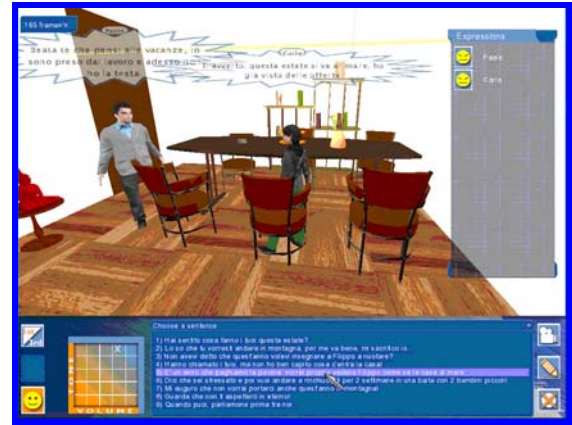- more Conciliatory (+Con), Conciliatory (Con), less Conciliatory (-Con),

Fig. 1 – Negotiation session (two agents) from SISINE software

- more Neutral (+Neu), Neutral (Neu), less Neutral (-Neu),

- less Aggressive (-Agg), Aggressive (Agg), more Aggresive (+Agg).

These behaviors correspond to the actions such as: +Con (Action0), Con (Action1), -Con (Action2), +Neu (Action3), Neu (Action4), -Neu (Action5), -Agg (Action6), Agg (Action7), and +Agg (Action8).

Note that, in this work, + Neu is considered belonging to positive socio-emotional area than - Neu.

The answers are linked to actions such as for instance, the answer corresponding to +Con(Action0), is considered as a more conciliatory answer, or the answer corresponding to Con(Action1), is considered as a conciliatory answer, and so on.

Then, after learning, the answer corresponding to the character with the max Qvalue is chosen. Rewards given in Table 1 for the suggested basic buying strategy, are used for training. Note that, for most part of everyday negotiations, a completely conciliatory agent strategy, or completely neutral agent strategy, or completely aggressive agent strategy are insufficient. Then, an interesting strategy is necessarily a combination of these three basic conciliatory, neutral, and aggressive strategies. Such interesting combination is given through a basic buying strategy which could be neutral (+Neu) for some first states, then aggressive (-Agg) for some second states, and conciliatory (-Con) for the last states, suggesting for buying something with less price, it is interesting to negotiate at first neutrally, then aggressively, and finally conciliatory to conclude (buy).

The technology of multi-agent systems which facilitates the negotiation at operative level of the decision-making [18] is used. It allows agents to embody a notion of autonomy, in particular, to decide for themselves whether or not to perform an action on request from another agent. More, in order

to satisfy their design objectives, agents are designed to be intelligent, i.e., capable of flexible behavior [1], [7], [8]: able to perceive their environment, and respond in a timely fashion to changes that occur in it (reactivity), able to exhibit goal-directed behavior by taking the initiative (proactiveness), and capable of interacting with other agents and possibly humans (social ability) ; such abilities are guaranteed in this work by psychological aspects of the personality, multi-agents, and Q-Learning.

Table 1. Strategy rewards (basic buying strategy)

	States From S0 to S8				States From S9 to S16			
+ Con	- 60	- 60	...	- 60	- 60	- 60	...	- 60
Con	- 50	- 50	...	- 50	- 50	- 50	...	- 50
- Con	- 40	- 40	...	- 40	- 40	- 40	...	- 40
+ Neu	**+ 1**	**+ 1**	**...**	**+ 1**	- 30	- 30	...	- 30
Neu	- 1	- 1	...	- 1	- 20	- 20	...	- 20
- Neu	- 2	- 2	...	- 2	- 10	- 10	...	- 10
- Agg	- 40	- 40	...	- 40	**+ 1**	**+ 1**	**...**	**+ 1**
Agg	- 50	- 50	...	- 50	- 1	- 1	...	- 1
+ Agg	- 60	- 60	...	- 60	- 2	- 2	...	- 2

	States From S17 to S24			
+ Con	- 2	- 2	...	- 2
Con	- 1	- 1	...	- 1
- Con	**+ 1**	**+ 1**	**...**	**+ 1**
+ Neu	- 10	- 10	...	- 10
Neu	- 20	- 20	...	- 20
- Neu	- 30	- 30	...	- 30
- Agg	- 40	- 40	...	- 40
Agg	- 50	- 50	...	- 50
+ Agg	- 60	- 60	...	- 60

In order to elaborate such negotiation strategy, SOcial and COgnitive SYStem (SOCOSYS) which is presented in Fig. 2, is developed. It is built of three main parts: a multi-agent system representing the environment model, an intelligent agent, and the simulation environment (SISINE software). Thus, SOCOSYS is designed and developed as a social and cognitive system for learning negotiation strategies (human-agent and agent-agent).

4. NEGOTIATION BEHAVIORS

One of the most important breakthroughs in reinforcement learning was the development of an off-policy temporal-difference control algorithm known as Q-learning and Sarsa-learning [10], [11], [12]. Elsewhere, neural network (fuzzy artmap neural network in this work) implementation of reinforcement Q-learning offers the advantages of learning and generalization (essential traits of intelligent behaviors), robustness, massively parallel computations and distributed memory [14], [15], [16] and limited memory requirement for storing the knowledge [19], [20].
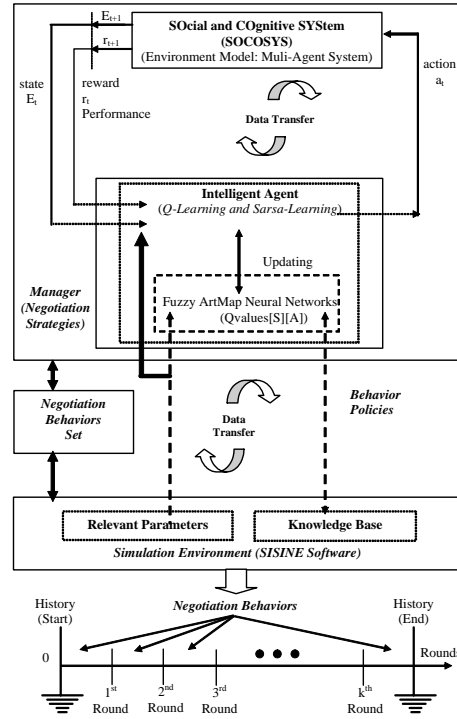
Fig. 2 – SOcial and COgnitive SYStem (SOCOSYS) for learning negotiation strategies (human-agent and agent-agent)

4. 1. REINFORCEMENT LEARNING

Reinforcement learning allows an agent (the learner and decision-maker) to use its experience, from the interaction with an environment, to improve its performance over time [10], [11], [12]. In other words, from the interaction with an environment an agent, can learn, using reinforcement Q-learning, to maximize the reward *r* leading to an optimal behavior policy. Indeed, in this on-line reinforcement learning, the agent incrementally learns an action/value function Q(s, a) that it uses to evaluate the utility of performing action *a* while in state *s*. Q-learning leads to optimal behavior, i.e., behavior that maximizes the overall utility for the agent in a particular task environment [7]. The used Q-learning [10], [12] and Sarsa-learning [12] paradigms have been detailed in [17].

The parameter settings of the initial Q values, constant step-size parameter ($0 <= \alpha < 1$), and discount rate ($0 <= \gamma < 1$) have been done following the choice approaches given in [10] and [12] resulting in: initial Q values = 0.5, $\alpha = 0.1$, and $\gamma = 0.01$. The results of Q-Learning of the basic buying strategy, from the strategy rewards given in Table 1, with $\gamma = 0.01$ are globally following the rewards. Q-Learning succeeded in learning the strategy giving the maxQvalues (1.0) for Action3 in the states (or rounds) from S0 to S8, then giving the maxQvalues (1.0) for Action6 in the states from S9 to S16, and finally giving the maxQvalues (1.0) for Action2 in

the states from S17 to S24, as shown in Fig. 3. Note that similar results of Sarsa-Learning (basic buying strategy), from the strategy rewards given in Table 1, are obtained with $\gamma = 0.01$.
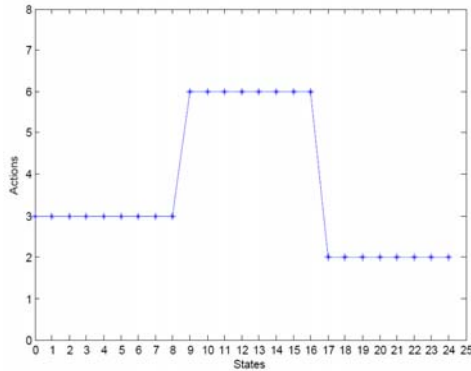

Fig. 3 – Strategy results (Q-Learning: $\gamma = 0.01$)

For other values of γ, the Sarsa-Learning offers the advantage of safe solutions (not optimal solutions which could be dangerous offered by Q-Learning) [21], [22]. Indeed, in this case, Q-Learning offers optimal solution (similarly to the rewards shown in Table I) corresponding to Action3. However, in on-line learning some results could be dangerous in the sense that, when learning Action3 (+Neu), the result could be related sometimes to Action2 (-Con) or Action4 (Neu). In the case of Action2 (-Con), this is dangerous because of the complete changing of the character under learning from neutral character to conciliatory character. Thus, Sarsa-Learning is offering more desirable properties to the suggested character learning than Q-Learning.

In this work, such safe solution (Sarsa-Learning) is obtained with $\gamma = 0.86$, as detailed in [17] and shown in Fig. 4. In fact, for state 0 to 8, action under learning is Action3 (+Neu) an offered solutions are Action4 (Neu) and Action5 (-Neu) ; for state 9 to 16, action under learning is Action6 (-Agg) an the offered solutions are Action6 (-Agg) and Action7 (Agg) ; and finally for state 17 to 24, the action under learning is Action2 (-Con) an the offered solutions are Action1 (Con) and Action2 (-Con).
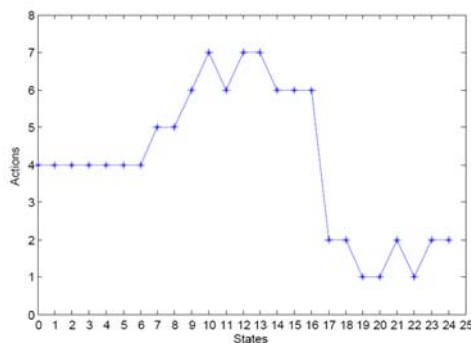

Fig. 4 – Strategy results (Sarsa-Learning: $\gamma = 0.86$)

4. 2. FUZZY ARTMAP NEURAL NETWORK (FAMNN)

One of the main features of human memory is its ability to learn many new things without necessarily forgetting things learned in the past. The ability of humans to remember many details of an exciting movie is a typical example of fast learning. Such learning can be achieved exploiting the Adaptive Resonance Theory (ART) in combination with neural networks and fuzzy logic. Indeed, ART can be used to design a hierarchical artmap neural networks that can rapidly self-organize stable categorical mappings between i-dimensional input vectors and n-dimensional output vectors [23]. To better reflect human reasoning, fuzzy inference is incorporated to the basic architecture leading to Fuzzy ArtMap Neural Networks (FAMNN) [15], [23]. FAMNN are capable of *fast* and *stable* learning of recognition categories in response to arbitrary sequences of input patterns. Therefore, they achieve a synthesis of fuzzy logic and ART neural networks by exploiting a close formal similarity between the computations of fuzzy subset hood and ART category choice, resonance, and learning.

FAMNN classifier is then trained to acquire the suggested strategy (a basic buying strategy), from the learning algorithm detailed in [23], from one hundred (100) examples of the training set (4x9 examples for Action3, 4x8 examples for Action6, and 4x8 examples for Action2) with a normalization of the inputs (between 0 and 1) either by incorporating an example into an existing output category node or creating a new output category node for it. This classifier sprouted the output category node number OCNN = N with N = 9 and yields convergence in well under the cycle number CN = 1 with the learning rate $\eta = 1$, small positive constant $\lambda = 0.000001$, baseline of the vigilance $\sigma = 0.4$. FAMNN results demonstrate a fast and stable learning (all adaptive weights only decrease in time).

After learning, FAMNN classifier is built of raw input layer (complement coder), input layer, output category layer, and category layer shown in Fig. 5.

Raw Input Layer: with three inputs corresponding to components of the input vector **X**: strategy (X_0), character (X_1), and state or round (X_2).

Input Layer: with six input nodes, constituting the input vector **I** in Eq. (1), obtained from the three inputs implemented in a complementary coding form by a complement coder from an input i-dimensional vector to 2i dimensional vector.

$$\mathbf{I} = [X_1, X_2, X_3, \overline{X}_1, \overline{X}_2, \overline{X}_3]. \qquad (1)$$

Output Category Layer: starts with no output category nodes (no weights) and grows reaching finally nine output nodes.

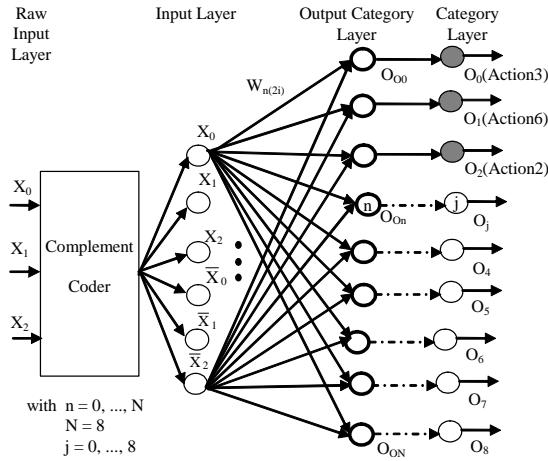**Fig. 5 – Fuzzy Artmap Neural Network Architecture**

For each input **I**, the output activation function (choice function), T_n, is defined by Eq. (2):

$$T_n(\mathbf{I}) = \frac{\left|\mathbf{I}_{(2i)} \wedge \mathbf{W}_{n(2i)}\right|}{\lambda + \left|\mathbf{W}_{n(2i)}\right|}, \qquad (2)$$

where λ is a small positive constant $0 < \lambda \ll 1$.

Fuzzy AND operator $\wedge$ is defined by Eq. (3):

$$(\mathbf{p} \wedge \mathbf{q})_m \equiv \mathrm{Min}(p_m, q_m), \qquad (3)$$

with $m = 1, \ldots, M$; norm $| \, . \, |$ is defined by Eq. (4):

$$|\mathbf{p}| \equiv \sum_{m=1}^{M} |p_m|, \qquad (4)$$

for any m-dimensional vectors **p** and **q**.

Then, category choice in Eq. (5) is indexed by N:

$$T_N = \mathrm{Max} \{ T_n: n = 1, \ldots, N \}. \qquad (5)$$

Category Layer: with nine category nodes O_j, the maximum number of categories the network can learn, each to be labeled as a unique category (class).

Then, the obtained weights are used (stored) in ROMs of the following FPGA architecture.

5. FIELD PROGRAMMABLE GATE ARRAY (FPGA) ARCHITECTURE

From the design with discrete components containing simple gates, the industry has moved from TTL to CMOS, then PLDs and now FPGAs and ASICs. Each of these moves has contributed to higher operating speeds, lower power consumption, and easier design as well as reduction in physical size. However, several factors, affect the decision to implement a design using an ASIC or an FPGA. For instance, current ASICs have a low cost-per-gate advantages as well as an inherent speed advantage ; in contrast, FPGAs have been winning with their time-to-application, low nonrecurring engineering fees, and reprogrammable features [24].

For this purpose, an FPGA architecture is suggested for FAMNN integration, where neurons

are modeled using Weights, MIN operator and Norm and Eq. (2) as shown in Fig. 6.
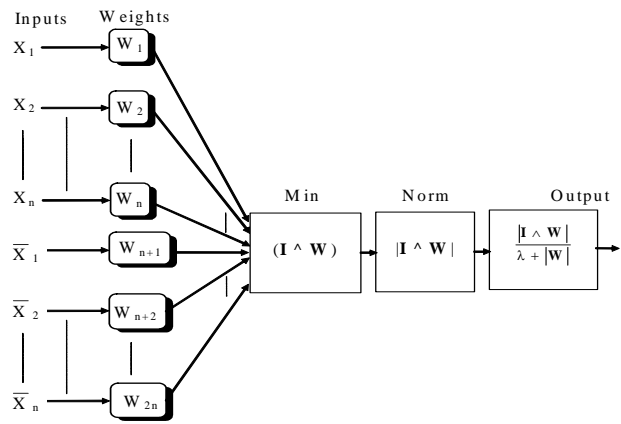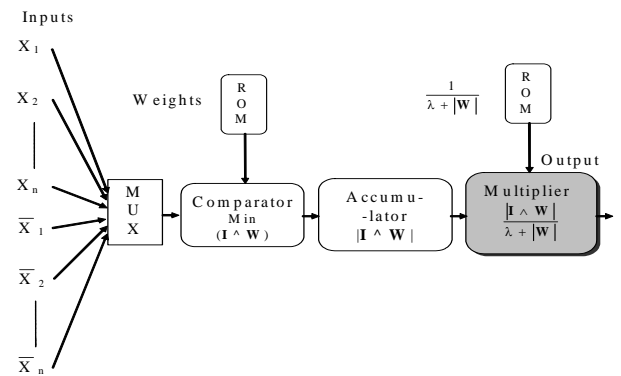

Fig. 6 – Neuron model
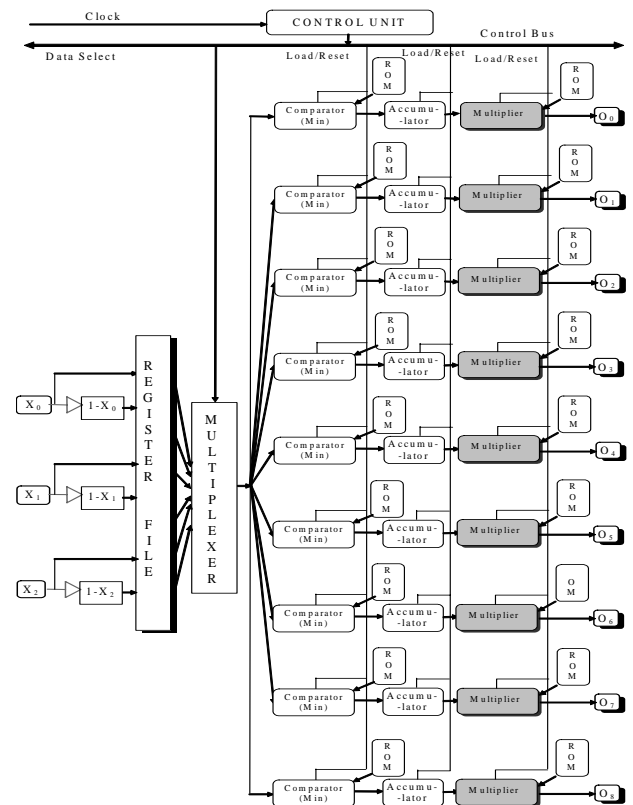

Fig. 7 – Neuron model hardware

Fig. 8 – FPGA architecture for FAMNN integration

Then, the corresponding hardware of the neuron

model uses multiplexers (MUX), comparators, accumulators, multipliers, and ROM memories where the obtained weights (after learning) are stored as well as the corresponding term $1/(\lambda+|\mathbf{W}|)$, see Eq. (2)., as shown in Fig. 7.

Thus, the FPGA architecture for FAMNN integration is suggested in Fig. 8. In this architecture, the result of the complement coder is stored in a register file, then a multiplexer matches each input to each neuron. Note that such architecture is simpler and uses less hardware than multilayer feedforward neural networks, fuzzy neural networks, or fuzzy systems [25].

6. CONCLUSION

In this paper, a negotiation strategy essentially based on the psychological aspects of the human personality (negotiation behaviors: +Con, Con, -Con, +Neu, Neu, -Neu, -Agg, Agg, +Agg) is suggested for SISINE. Such negotiation behaviors are acquired by reinforcement Q-learning and Sarsa-Learning approaches in order to guarantee the agent reactivity, agent proactiveness, and agent social ability. The analysis and comparison show that Sarsa-Learning offer the advantage of safe solutions with regard to Q-Learning. The strategy is then acquired by FAMNN exploiting particularly their fast and stable learning capabilities (essential trait of intelligent behavior), massively parallel computations, and limited memory requirement for storing the knowledge. Moreover, an FPGA integration of FAMNN has been suggested, which offers a simpler and uses less hardware than multilayer feedforward neural networks, fuzzy neural networks, or fuzzy systems [25]. Thus, the suggested strategy displays the ability to provide agents, through a basic buying strategy, with a first intelligence level in SOCOSYS for learning negotiation strategies (human-agent and agent-agent).

An interesting alternative for future research is the investigation with regard to the results of the suggested strategy based on the proposed behaviors which should be enhanced from a buying strategy acquired by learning from a human expert buyer.

7. REFERENCES

[1] M. Wooldridge, *An Introduction to MultiAgent Systems*. John Wiley & Sons, England, 2002.

[2] R. F. Bales, *Interaction Process Analysis: A Method for the Study of Small Groups.* Cambridge, Massach.: Addisson-Wesley, 1950.

[3] J. Z. Rubin and B. R. Brown, *The Social Psychology of Bargaining and Negotiation.* NY Acad. Press, pp. 233-262, 1975.

[4] D. Silver, R. Sutton, and M. Müller, "Reinforcement Learning of Local Shape in the Game of Go", *Int. Joint Conference on Artificial Intelligence*, 2007, pp. 1053-1058.

[5] O. Miglino, A. Di Ferdinando, A. Rega, and B. Benincasa, "SISINE: Teaching negotiation through a multiplayer online role playing game", *The 6th European Conference on E-Learning*, Danemark, Oct. 04-05, 2007.

[6] Orazio Miglino, "The SISINE Project: Developing and E-Learning Platform for Educational Role-Playing Games", Technology-Enhanced Learning, *European Research Consortium for Informatics and Mathematics* (71) (2007), pp. 28-29.

[7] R. Pfeifer and C. Scheier. *Understanding Intelligence.* MIT press, 1999.

[8] A. Chohra, *Embodied Cognitive Science, Intelligent Behavior Control, Machine Learning, Soft Computing, and FPGA Integration: Towards Fast, Cooperative and Adversarial Robot Team (RoboCup).* Technical GMD Report, No. 136, Germany, June 2001.

[9] D. Zeng and K. Sycara, "Benefits of learning in negotiation", *Proc. of the 14th National Conference on Artificial Intelligence (AAAI-97)*, Providence, RI, July 1997.

[10] C. J. C. H. Watkins. *Learning from Delayed Rewards.* Ph.D Thesis, King's College, 1989.

[11] S. D. Whitehead. *Reinforcement Learning for the Adaptive Control of Perception and Action.* Technical Report 406, Rochester University, 1992.

[12] R. S. Sutton and A. G. Barto. *Reinforcement Learning.* MIT press, 1998.

[13] T. G. Dietterich, "Hierarchical reinforcement learning with the MAXQ value function decomposition". *Journal of Artificial Intelligence Research* 13 (2000), pp. 227-303.

[14] J. A. Anderson, *An Introduction to Neural Networks*, MIT Press, England, 1995.

[15] D. W. Patterson, *Artificial Neural Networks: Theory and Applications*, Prentice-Hall, Simon & Schuster, Singapore, 1996.

[16] S. Haykin, *Neural Networks: A Comprehensive Foundation.* 2nd Ed., Prentice-Hall, 1999.

[17] Amine Chohra, Arash Bahrammirzaee, Kurosh Madani, "Negotiation Agent Behaviors Based On Reinforcement Learning Approaches and Fuzzy ArtMap Neural Networks", Fifth Int. Conf. on Neural Networks and Artificial Intelligence (ICNNAI 2008), Minsk, Belarus, May 27-30, 2008, pp. 56-61.

[18] T. W. Sandholm, *Distributed Rational Decision Making.* In Multiagent Systems: A Modern Introduction to Distributed Artificial Intelligence, MIT Press, pp. 201-258, 1999.

[19] L.-J. Lin, "Self-improving reactive agents

based on reinforcement learning, planning and teaching'', *Machine Learning*, 8 (1992) pp. 293-321.

[20] C. F. Touzet, ''Neural reinforcement learning for behaviour synthesis'', *Robotics and Autonomous Sys.* 22 (1997), pp. 251-281.

[21] M. Coggan, "Exploration and Exploitation in Reinforcement Learning", Research supervised by Prof. Doina Precup, CRA-W DMP Project at McGill University, 2004.

[22] K. Takadama and H. Fujita, "Lessons learned from comparison between Q-learning and Sarsa agents in bargaining game", North American Asso. for Computational Social and Organizational Science Conference 2004.

[23] G. A. Carpenter, S. Grossberg and D. B. Rosen, ''Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system'', *Neural Networks* 4 (1991), pp. 759–771.

[24] R.K. Awalt, 'Making ASIC/FPGA design'', *Integrated System Design* (1999) pp. 22-28.

[25] O. Azouaoui, A. Chohra, ''Soft computing based pattern classifiers for the obstacle avoidance behavior of Intelligent Autonomous Vehicles (IAV)'', *Int. Journal of Applied Intelligence* 16 (3) (2002), pp. 249-271.

Mr. Arash Bahrammirzaee Received his B.S in Insurance Management in August 2005 from Eco College of Insurance of Allameh Tabatabaiy University at Tehran (Iran). After two years (in September 2007) he received his Master degree in Information Technology Management from Tarbiat Modares University at Tehran (Iran). Currently, he prepares a PhD in Computer Science at Images, Signals and Intelligent Systems Laboratory (LISSI / EA 3956) of PARIS XII University, Lieusaint (France) since November 2007. His research interests are negotiation behaviors, Artificial Intelligence, artificial neural networks, expert systems, and decision support systems.

Prof. Kurosh Madani Received his Ph.D. degree in Electrical Engineering and Computer Sciences from University PARIS XI, Orsay, France, in 1990. From 1989 to 1990, he worked as assistant professor at Institute of Fundamental Electronics of PARIS XI University. In 1990, he joined Creteil-Senart Institute of Technology of University PARIS XII – Val de Marne, Lieusaint, France, where he worked from 1990 to 1998 as assistant professor. In 1995, he received the DHDR Doctor Habilitate degree (senior research Dr. Hab. degree) from University PARIS XII – Val de Marne. Since 1998 he works as Chair Professor in Electrical Engineering of Senart Institute of Technology of University PARIS XII. From 1992 to 2004 he has been head of Intelligence in Instrumentation and Systems Laboratory (I2S / JE 2353). Since 2005, he is head of one of the three research groups of Images, Signals and Intelligent Systems Laboratory (LISSI / EA 3956) of PARIS XII University. He has worked on both digital and analog implementation of processors arrays for image processing, electrooptical random number generation, and both analog and digital ANN implementation. His current research interests include large ANN structures modeling and implementation, hybrid neural based information processing systems and their software and hardware implementations, design and implementation of real-time neurocontrol and neural based fault detection and diagnosis systems. Since 1996 he is a permanent member (elected Academician) of International Informatization Academy. In 1997, he was also elected as Academician of International Academy of Technological Cybernetics.

Dr. Amine Chohra Received his Doctorate esssciences in 1999 from Ecole Nationale Polytechnique, Algiers (Algeria). He has been a member of Artificial Intelligence and Robotics Lab. (LRIA) of CDTA from 1991 to 1999. From 1999 to 2001, he worked as post doctoral researcher, with Behavior Engineering team of AiS-GMD, Sankt Augustin (Germany) and with Dependable Computing Group of IEI-CNR, Pisa (Italy), respectively. From 2001 2003, he has been teacher/researcher at Orleans University, ENSI de Bourges (France), and a member of Vision and Robotics Laboratory (LVR / UPRES EA 2078). Since September 2003, he is Assistant Professor at Senart Institute of Technology of PARIS XII University, Lieusaint (France) and a staff member of Images, Signals, and Intelligent Systems Laboratory (LISSI / EA 3956) of this University. His research interests are information processing systems, knowledge based systems, hybrid intelligent systems, soft computing, machine learning and decision-making, pattern recognition, social and cognitive systems, and computer aided diagnosis.