Elaboration of Cognitive Decision Making Methods in the Context of Symbiotic Networking

Milos Rovcanin, Eli De Poorter, Opher Yaron, Ingrid Moerman, David Plets, Wout Joseph, Luc Martens

Ghent University - IBBT, Department of Information Technology (INTEC), IBCN-WiCa

Gaston Crommenlaan 8, Bus 201, 9050 Ghent, Belgium

milos.rovcanin@intec.ugent.be, {*firstname.lastname*}@intec.ugent.be

Abstract—Recently, the concept of 'cognitive networking' has been introduced, in which reconfigurable radio networks rely on self-awareness and artificial intelligence to optimize their network performance. These cognitive networks are able to perceive current network conditions and then plan, learn and act according to end-to-end goals. This paper elaborates on different methods (network solutions) that can be used by cognitive networks for deciding on how to optimize the performance of a large number of co-located devices with different characteristics and network requirements. To this end, a negotiation based networking methodology ('symbiotic networking') is used that supports efficient network cooperation between heterogeneous devices in order to optimize their network performance. In this paper, the advantages and disadvantages of different reasoning techniques that can be used during the decision making phase are discussed.

Index Terms—symbiotic cognitive networks; reasoning methods; machine learning; game theory;

I. INTRODUCTION

Wireless networks are becoming increasingly complex, heterogeneous and dynamic, which motivate the evolution of the concept of cognitive networks. As it is described in [1]: "A wireless cognitive network is a network with a cognitive process that can perceive current network conditions and then plan, decide and act on those conditions". The network can learn from these adaptations and use them to make future decisions, all while taking into account end-to-end goals.

Cognition process (Fig.1), in this case, is related to machine learning [2] as any algorithm that: "improves its performance through experience gained over a period of time without the complete information about the environment in which it operates". This process can be divided into four stages:

(i) *Gather observations* of the important aspects (properties) of the network or, in other words, gather knowledge. Inside complex systems, with a large number of nodes, it is more likely that the cognitive process is performed with an incomplete knowledge about the system status. (ii) *Plan actions* according to the network policies and the knowledge that was gathered. Reasoning is used to decide which scenario best fits the end-to-end goals. After the decision is made, (iii) actions are performed accordingly. The (iv) cognitive feedback loop measures the success of the chosen solution relative to the defined objectives. This way, when similar circumstances happen in the future, the cognitive decision maker will have an idea what decisions are preferred and which ones should be avoided.
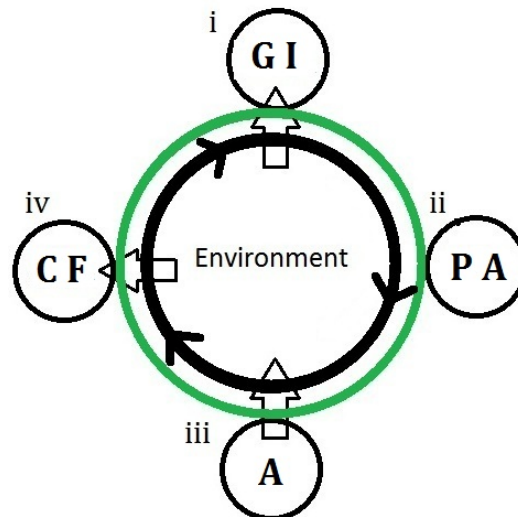

Fig. 1. A four stage cognition cycle: (1) data gathering (DG), (2) planning (PA), (3) acting (A), (4) collecting feedback (CF)

The fundamental difference between a cognitive network and a cognitive radio are the end-to-end goals. They give a cognitive network its network-wide scope, separating it from other adaptation approaches, which usually have only a local, single element scope [3].

Cognitive networking solutions can be applied to improve network performance in situations in which different networks cooperate in an ad-hoc and dynamic way. For example, co-located IEEE802.11 WiFi, IEEE802.15.4 sensor and 802.15.1 Bluetooth networks are typically configured independently from each other: their settings take into account only the behaviour of their own communication technology and they typically ignore their influence on each other. By ensuring that different co-located networks are aware of each other, they can modify their configuration so that the performance of all individual networks improves. The cognition loop can be used as the way to further improve the newly formed symbiotic network performance. Some of the examples are: improving reliability, decreasing energy consumption, lowering exposure etc.

The remainder of the paper is organised as follows: The SymbioNets use case, a concept of symbiotic wireless sensor networks cooperation, is presented in Section II. Section III brings the overview of the most commonly used cognitive

methods in the context of wireless (sensor) networking. In Section IV we compare suitability of the above mentioned approaches, from several different aspects, for the symbiotic networking. Conclusion is given in Section V.

II. THE SYMBIONETS USE CASE

The need for cognitive networking is also stated in the SymbioNets project [4]. In this project, different networks engage in cooperation by activating specific symbiotic network services (Fig.2), when activating these services result in better network performance for all involved networks. These symbiotic network services are not crucial for the correct operation of the individual networks, but instead influence the behaviour of the communication.
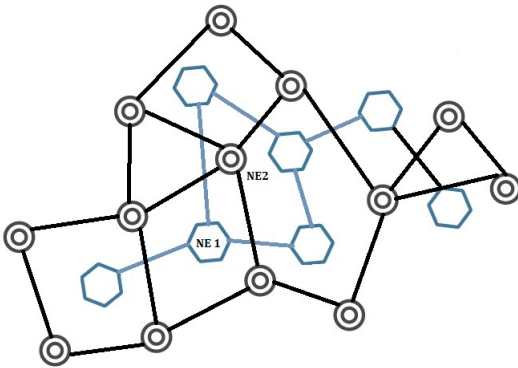

Fig. 2. Coexisting networks suitable for symbiotic cooperation with their negotiation representatives (NE). An example would be: temperature monitoring network (black nodes) and a security network (blue hexagons) inside a facility

By activating a symbiotic network service, the reliability, energy consumption, exposure, etc. of one or more networks is influenced. In the SymbioNets project, a number of symbiotic services is dynamically activated or deactivated based on the network requirements. Some example symbiotic network services are the following:

- A network can offer Internet access to other networks
- An 'interference avoidance' algorithm can be activated to reduce interference
- A 'packet sharing service' allows cooperating networks to interpret and route packets from each other

The full SymbioNets cognitive cycle works as follows. Co-located communities of devices exchange their service profiles so that each one of them has an idea about the network preferences of co-located devices. These preferences describe behavioural aspects of the network (e.g., 'limit the battery consumption') or express the need for additional functionality (e.g., 'get Internet access'). There is a negotiation entity (NE) in every community which collects all service profiles of neighbouring communities and decides which symbiotic network services should be activated or deactivated. The process

of negotiation is rerun every time a new or an updated profile is received. Service negotiation messages are exchanged between negotiation entities to reach a common decision. Then, the decision is disseminated to every node in the community.

One critical aspect of the SymbioNets approach is that the networks need to decide which network services to activate based not only on environment observations, but also in networks consisting of strongly heterogeneous devices with diverging communication technologies, diverging application requirements and diverging computational resources. In the current implementation, a linear programming algorithm, that is implemented on the negotiation node [4], is used to automatically calculate the optimal set of network services to be enabled on every node. However, this solution assumes that benefits and costs of enabling each service are predefined and fixed. In a dynamic environment this assumption usually does not hold. Instead, a more advanced cognitive decision method is required. In this paper, we describe a number of cognitive decision methods that can be used and explain the concepts on which they rely. Most importantly, we point out their major advantages and drawbacks if applied in heterogeneous, symbiotic environments. These cognitive methods are evaluated on several criteria:

- *Complexity*. Are they also suitable for use on devices with low processing power, such as sensor nodes?
- *Support for heterogeneity*. Suitable for an environment where different communication technologies and different types of application requirements coexist.
- *Dealing with "'malicious"' behaviour* in the network. For example, what if certain devices try to 'cheat' by falsely reporting that they activated certain services?
- *Support for distributed solutions*. Suitability for networks where devices have to make decisions based on local information, rather than having a full network overview.

III. COGNITIVE METHODS

This section gives an overview of relevant decision making methods and elaborates on their advantages and suitability for heterogeneous, symbiotic networks. We discuss the following decision methods: mathematical methods (including linear programming), interactive network simulators, game theory and machine learning.

A. Mathematical methods

In small networks, it may possible to calculate the influence of a specific network configuration on the network performance. As an example, consider *queuing theory*, that can be used to derive performance metrics such as the average waiting time in queues of the system, the expected number of packets in queues, packet drop probabilities, etc. However, queuing theory is not capable of predicting end-to-end network qualities in large-scale and complex networks that use multiple network protocols.

Other mathematical models were developed to calculate the performance of higher-layer network protocols. For example,

in [5] a mathematical model is created to optimize the stability and fairness of rate control algorithms for the Internet. However, these formulas require in-depth knowledge of the innerworking of the specific network protocols.

Finally, in [4], a linear program is used to calculate which, among a list of potential network functions, should be activated to best suit the network requirements of several co-located networks. For example, if energy efficiency of a smartphone device is an issue and no low-latency applications are being used, the linear program can activate a packet aggregation protocol. For its calculations, the linear program assumes that the influence of network protocols is known in advance.

All these mathematical methods have the same disadvantages. (i) They do not accurately model the whole networks, or use abstractions of certain parts of the network. (ii) They are not suitable for modeling complex cross-layer and cross-network interactions. (iii) And finally, they assume perfect knowledge about the network. On the other hand, mathematical formulas often have a low processing overhead and they can accurately predict a number of key performance metrics.

B. Network simulators

As an alternative, it is possible to use a *network simulator* in the decision making process. The cognitive engine can recreate (to the best of its abilities) the monitored network conditions in a network simulator, and use the simulation to predict the influence of optimization decisions. Existing network simulators such as ns2/ns3, OPNET, Netsim, etc. have become increasingly accurate and are capable of taking into account many cross-layer and even physical layer interactions. A similar approach is the use of *planning tools* for decision making. Planning tools are typically used before deployment of a network to predict the behaviour of (wireless) networks in different environments [6]. They are, for example, used to calculate the optimum number of Wi-Fi access points and communication settings to obtain a certain network performance.

A cognitive reasoning engine can be created by recreating the network set-up in a network simulator or planning tool, and running multiple scenarios with different settings. This way, the decision engine can estimate and select the best network configuration.

The main disadvantages with this approach are the following. (i) Network simulators are typically not designed for distributed calculations. (ii) Detailed information about the network is required, such as the location of each device, the exact network configuration and settings and the application requirements. (iii) Current network simulators do not take into account the influence of different co-located communication technologies that use the same frequency band. Finally, (iv) network simulators often require heavy processing power.

C. Game theory

Game theory models the behaviour of a system as a game, played by at least two rational players. Rationality, in this context, means balancing costs against benefits to arrive at an action that maximizes personal payout. Game theory has been used in logic, economics, psychology, political science, biology etc. In the domain of computer networks, it offers tools that can be used in modelling an interaction among individual nodes in a wireless network. Game theory enables us to determine the existence, uniqueness and convergence to a steady state operating point.

Formally, a game is given by $G = (N, A, u_i)$, where $N = 1, 2, ..., n$ is the set of players, A_i is the action set for player i. u_i is the set of utility functions that each player wishes to maximize, where $u_i : A_i \rightarrow \Re$. For every player, i, the utility function is a function of the action chosen, a_i and the actions chosen by all the players in the game other than player i, denoted as a_{-i}. Together a_i and a_{-i} make up the action tuple a. An action tuple is a unique choice of actions by each player.

Steady-state conditions, the Nash equilibrium, can be identified using this model. The Nash equilibrium represents a state where an individual node cannot increase the values of its utility function by changing actions, assuming other players remain constant in their strategies. Pareto efficiency, on the other hand, describes how acceptable is the achieved steady state from a global point of view. A Pareto optimal state is the one where no player can benefit from any action without making someone else worse off.

Problems such as node's power consumption, contention for a communication medium and routing in cognitive wireless networks have already been presented in the form of a game [7].

For our specific use case, described in Section II, the negotiation process should include a process of defining pricing policies and action sets, taking into account the wireless technology each community is based on. The disseminated list of enabled services defines the starting point in a game. Nodes aim to increase their utility functions regarding each given incentive and a properly designed pricing policy will lead to an increase of the Pareto efficiency. By sending feedback to the negotiation node, service policies can be updated to trigger recalculation and renegotiation. Since every network preference that needs to be optimized in the community demands a distinctive game, it is likely that a number of these different games will involve common parameters. Certain trade-offs can be foreseen. For example: a packet sharing and a power saving game can be played simultaneously. While the first one demands a node to forward more packets thus spend more energy, the second one will try to minimize the power consumption as much as possible. There has to be a trade-off between spending less energy per node and ensuring better network coverage and shorter routing paths.

One difficulty of using the game theory in SymbioNets, is how to correctly interpret low-level actions into high-level goals. Additionally, the set of available actions to the players needs to be carefully defined and verified as well as the states the system can go through. Finally, computational complexity will increase significantly as the number of nodes grows.

D. Machine learning

Machine learning is a form of artificial intelligence in which devices learn using inductive inference (Fig.3) [8] [9]. One specific field, *Reinforcement learning*, is particularly suitable for usage in the context of cognitive networks [10]. It involves the notion of learning through trial-and-error interaction with a dynamic environment.

This concept can be described as follows: a decision making node has a set of possible actions $A = a_1, a_2, ..., a_N$ at each step. Based on observations of the environment, the next step is chosen. It is desirable to have a process with a finite number of states. When the action is taken, the environment makes a transition to another state according to a certain probability distribution $P(s'|s, a)$ and so on. The main goal of a decision maker is to maximize its action-value function (Bellman's equation) :

$$Q(s, a) = r(s, a) + \gamma \sum_{s'} P(s'|s, a) maxQ(s', a')$$

The Q function assigns a value to each state/action pair. This value is increased every time a decision maker takes action a at state s. It also takes into account the highest available Q value at each connected state. The factor γ makes sure the reward for making the same decision in the future decreases. $r(s, a)$ represents an immediate reward, the one that is awarded at the initial transition from the state s.
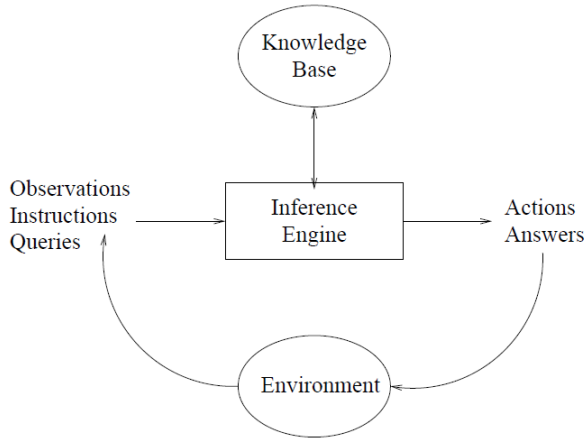

Fig. 3.   The standard model of a learning agent

In [11] a number of reinforcement learning (RL) technique applications in wireless ad-hoc and wireless sensor networks are presented. Reinforcement learning is described as very well suited for distributed problems, like routing. It has "medium" requirements for memory and computation at the individual nodes, is easy to implement, highly flexible to topology changes, but it needs some time to converge.

Q-learning algorithm [12] is probably the most frequently used technique of reinforcement learning in wireless ad-hoc networks. It does not need any model of the environment and

can be used for on-line learning of the value function of some RL task, referred to as the Q-function. It is relatively easy to implement and has a good balance of optimality to memory and energy requirements. In wireless sensor networks, it was mostly used for optimizing the routing algorithms [13].

LSPI (Least Square Policy Iteration) is a model-free algorithm, which calculates the Q value of every state as a linear combination of so called "weighted basis functions" [14].

$$Q(s, a, w) = \sum_k \phi_i(s, a) w_i$$

Each one of k basis functions $\phi_i(s, a)$ represents a certain information about each state-action pair (e.g., residual energy of s', link quality between s and s' etc.). Weights w_i are parameters of the linear equation that are learned by gathering samples $\langle s, a, r, s' \rangle$. Every sample describes the reward r received upon executing action a in state s, ending in state s'. The main advantages of this algorithms are: it converges more quickly than the Q-learning and it does not require carefully tuning initial learning parameters [15].

Additionally, in [16] the Collaborative Reinforcement Learning technique (CRL) is introduced as a model-based technique with collaborating RL agents. This algorithm is applicable in heterogeneous networks, where agents typically poses different capabilities. CRL allows newly discovered agents to negotiate the establishment of causally connected states with their neighbours by exchanging device capability information.

In the context of the SymbioNets use case, the negotiation process, described in Section II, is used to check the compatibility of neighbouring community's communication technologies. It aims at defining the initial values of all the configuration parameters regarding each enabled service in the community. These values are used as starting points in the cognition process. A periodic feedback (values of a predefined set of parameters) are sent to a negotiation entity. After investigating the progress that has been made, the negotiation node decides whether to renegotiate the service policy or not.

The process of learning should be applied to every established network preference. Since the best possible action at some state is never known a priori, the decision making node needs to try different actions and sequences of actions so it could learn from its experiences. The process of maximizing action-value functions assigned to each preference enables the node to discover an optimal action-policy. This policy maps input values (gathered from the environment) into actions. As mentioned in the previous section, multiple processes are performed in parallel, including some common parameters which might lead to conflicting situations. It is necessary to determine correct trade-offs. This analysis can be done during the negotiation phase.

One of the main issues with reinforcement learning is that the number of possible environment states is large. To calculate the reward of a certain action, a node has to, at least, take a good guess what state the system is going to move into by preforming an action. The other problem is how to gather all

the necessary observations for multi-objective learning. Should there be a predefined set of data that is to be gathered at each step or should it depend on objective? Another pitfall is that, since the number of nodes will be fluctuating, it will be hard to establish a solid and non-changing policy since the influence of each action could change as the network topology changes.

IV. Choosing the optimal reasoning engine

This section compares the suitability of different reasoning engines for the symbiotic networking paradigm that was presented in Section II.

A. Supporting heterogeneity

Mathematical formulas typically assume that the network is very homogeneous. Heterogeneous networks might have diverging network requirements, different communication protocols and even use different communication technologies. Since new formulas have to be created for each different network situation, the use of mathematical formulas in not feasible for heterogeneous networks.

Current *network simulators* already have support for heterogeneity. They typically include a wide set of standardized network protocols and technologies, and as such can be used to estimate how different network settings influence performance. However, especially at the physical layer, a large number of unexpected cross-layer and cross-network interactions are typically simplified.

With regards to *game theory*, agents in heterogeneous networks can play a game, but with different perception of what goes on in the network, because the utility functions differ significantly - different power supply, radio transmission costs etc. This leads to different cost/benefit reasoning among nodes in different communities and has to be taken into account while negotiating optimal service and pricing policies.

Reinforcement learning algorithms support heterogeneity but every newly discovered agent has to negotiate the establishment of causally connected states with their neighbours by exchanging device capability information.

B. Complexity

The complexity of *mathematical formulas* depends strongly on the effect that is modelled (link local, end-to-end, ...) and the number of parameters that are taken into account. However, it is safe to say that highly complex networks can not fully be modelled using only mathematical formulas. In addition, solving complex mathematical formulas can be difficult on resource-constrained devices such as sensor nodes or smart phones.

The computational complexity of *network simulators* strongly depends on the accuracy of the simulator. Simulators that use approximations of communication effects can have low computational complexity, whereas extremely accurate simulators might require several days to calculate the behaviour of even small networks. However, even simplified network simulators are too complex to be used in resource-constrained networks: sensor networks or smart phones might want to off-load the cognitive reasoning process to a remote server that is less resource constrained.

In *game theory* approach, computation complexity increases as the number of players grows. Every individual player has to take into account actions of all other players. The existence of a steady state point must be proven before utilizing any specific algorithm. In the approach with the centralized authority (a negotiation node), the rules of a game can be changed during the play in respect of the feedback a negotiation entity is given from the network. Coordinated games assume exchange of information between nodes which can cost a lot in power consumption and memory space that is being used.

For *reinforcement learning*, model free algorithms need fresh data every once in a while, but the computational power and memory utilization are much lower than in model based algorithms. However, a model based algorithm can guarantee good results in an environment where acquiring measurements is highly expensive. With this approach, learning is faster, as much more use can be made of each experience. It can be used to solve multiple optimization problems. Many algorithms have already been implemented in WSNs, which implies that they can be adjusted to cope with severe power limitations. Complexity, however, does differ from algorithm to algorithm.

C. Dealing with "'malicious"' behaviour

In order to increase their own performance, devices can report false values, or fail to activate the network configuration settings they promised. A cognitive engine should be able to detect this kind of behaviour, and optionally punish these devices, or refuse to further cooperate with them.

When the predicted outcome of *mathematical formulas* differs from the measured performance, this can be an indication of 'cheating' nodes. However, it is almost impossible to detect the difference between incorrect predictions and malicious behaviour.

When the observed performance differs from the expected performance, *network simulators* can run through a large number of simulations to identify which settings produce similar behaviour. However, this process is likely very expensive computationally.

In *game theory*, a properly designed pricing policy aims to increase the Pareto efficiency of a game by making sure that every asocial behaviour is 'punished'. A node is considered to be malicious (selfish) if it tries to increase its own benefit without taking into account the social aspect of the game (making Pareto efficiency worse off).

When using *reinforcement learning*, Higher learning rates and not well defined rewarding policy will enforce particular states very quickly. In a dynamic network this can lead to suboptimal performance. One should trade-off between converging faster, but to a possibly mediocre operating point and converging at a slower rate while taking into account network changes and possible erroneous behaviour.

D. Suitability for distributed implementation

Mathematical methods typically require complete knowledge of the network configuration and as such can not be used

in a distributed way. The same is true for *network simulators*.

Game theory approach is naturally intended to operate in a distributed manner. Every node plays a game with one objective: to increase its payoff as much as possible. As mentioned before, to avoid selfish behaviour, the pricing policy must be properly defined. The existence of a supervising node, which gathers feedback information of all the players in a community and changes the rules and pricing policies on the fly, can be helpful, but fully centralized approach is inapplicable since it demands great computational power and full knowledge of the network.

Centralized *reinforcement learning algorithms* are inapplicable in resource-constrained wireless sensor networks since they assume a complete knowledge of the network's topology. If not impossible, in most cases this will be tremendously expensive to obtain. On the other hand, there already exist solutions for distributed reinforcement learning - Distributed Q-learning is a good example. MDQL is the solution for problems of multi-objective incentives in distributed manner [17].

V. Conclusion

This paper surveys different reasoning approaches and discusses how they can be applied to SymbioNets. Four different decision approaches are discussed. (i) *Mathematical approaches* require a low computational overhead, but are not well-suited to model complex cross-network and cross-layer influences. (ii) *Network simulators* can be used to determine optimal network settings by simulating a large number of network configurations. However, simulations have a large computational overhead and require perfect network knowledge at a central location. (iii) *Game theory* is well-suited for distributed negotiation implementations. However, each device needs an individual, custom designed cost and utility function. Finally, (iv) *machine learning approaches* do not require any knowledge about the innerworking of the network protocols. However, they can take a long time to reach a steady-state optimal network situation. As such, it is clear that no 'best' cognitive decision approach exists. Instead, the choice of approach depends on which SymbioNets criteria is deemed most important: *complexity*, *support for heterogeneity*, able to deal with *"'malicious"' behavior* and/or *support for distributed solutions*. Further work will focus on how to combine these different approaches to overcome their individual disadvantages.

Acknowledgment

References

[1] R. W. Thomas, L. A. DaSilva and A. B. MacKenzie, "Cognitive networks", Proc. IEEE DySPAN 2005, pp.352-60

[2] M. A. L. Thathachar and O. S. Sastry, "Networks of Learning Automata", Kluwer, 2004

[3] R. W. Thomas, D. H. Friend, L. A. DaSilva and A. B. MacKenzie, "Cognitive networks: Adaptation and Learning to Achieve End-to-End Performance Objectives", IEEE Communications magazine, Dec. 2006

[4] E. De Poorter, P. Becue, M. Rovcanin, I. Moerman and P. Demeester, "A negotiation-based networking methodology to enable cooperation across heterogeneous co-located networks", Ad Hoc Networks, Available online 8 December 2011, ISSN 1570-8705, 10.1016/j.adhoc.2011.11.007.

[5] F. Kelly, "Mathematical modelling of the Internet", Mathematics Unlimited - 2001 and Beyond, Springer-Verlag, Berlin, 2001. 685-702.

[6] D. Plets, W. Joseph, K. Vanhecke, E. Tanghe, L. Martens, Coverage Prediction and Optimization Algorithms for Indoor Environments, EURASIP Journal on Wireless Communications and Networking Special Issue on,Radio Propagation, Channel Modeling, and Wireless,Channel Simulation Tools for Heterogeneous, Networking Evaluation, accepted 2011.

[7] Jiahua Wu, "A Survey of Game Theory in Wireless Networking Design"

[8] T. G. Dietterich , Learning and reasoning, May 26, 2003

[9] A. Forster, "Machine Learning Techniques Applied to Wireless Ad-Hoc Networks: Guide and Survey", 3rd International Conference on Intelligent Sensors Sensor Networks and Information (ISSNIP), Melbourne, Australia, 3-6 Dec. 2007.

[10] T. G. Dietterich, and O. Langley, (2007) Machine Learning for Cognitive Networks:Technology Assessment and Research Challenges in Cognitive Networks: Towards Self Aware Networks, John Wiley and Sons, Ltd, Chichester,UK. doi: 10.1002/9780470515143.ch5

[11] L. P. Kaelbling, M. L. Littman, A. W.Moore, "Reinforcement learning: A Survey", Journal of Artificial Intelligence Research 4 (1996) 237-285

[12] C. Watkins, "Learning from delayed rewards,"Ph.D disseration, Cambridge university, Cambridge, England, 1989

[13] J. A. Boyan and M. Litman, "Packet routing in dynamically changing networks: A reinforcement learning approach," *Advances in Neural Information Processing Systems*, vol.6, 1994.

[14] M. Lagoudakis and R. Parr. Model-free least-squares policy iteration. In Proc. of NIPS, 2001.

[15] P. Wang, T. Wang, "Adaptive Routing for Sensor Networks using Reinforcement Learning", The Sixth IEEE International Conference on Computer and Information Technology, 2006. CIT '06.

[16] J. Dowling, E. Curran, R. Cunningham and W. Cahill, "Using feedback in collaborative reinforcement learning to adaptively optimize manet routing", *IEEE Transactions on Systems, Man and Cybernetics*, vol. 35, no. 3, pp. 360-372, 2005.

[17] C. E. Mariano, E. Morales, "A New Distributed Reinforcement Learning Algorithm for Multiple Objective Optimization Problems"