In Didactical Modeling Annals Vol 4:Year 2010/2011
Institute of Mathematics and Informatics, Bulgarian Academy of Science

TAXONOMY OF LEARNING AGENTS

Nevena Ackovska
Institute of Informatics,
FNSM, University St. Cyril and Methodius,
Skopje, Macedonia
nevena@ii.edu.mk

Abstract
This paper elaborates on taxonomy of learning agents and learning paradigms. A learning paradigm can be represented by an agent-environment interaction. Studying agent-environment interfaces is our approach toward understanding learning agents and learning systems in general. The paper observes ten learning paradigms ranging from context-free learning paradigms, to emotion based self-learning paradigms. The paper also discusses morphogenesis of learning agent architectures that allow agents to adapt and learn in various learning paradigms.

Keywords: behavioral environment, learning paradigms, genetic environment, emotion learning, morphogenesis of learning architectures

1. INTRODUCTION

The material presented here is part of our research in the Consequence Driven Systems (CDS) theory. Here we will present the approach, scope, and historical background before we elaborate further on learning paradigms. The CDS theory was proposed in 1981 as a solution of challenges within the theory of adaptive neural networks, particularly the delayed reinforcement learning problem.

A very brief history of the theory starts with initial mathematical description of functioning of a formal neuron (later named artificial neuron), as proposed by McCulloch and Pitts (1943). In 1958 Rosenblatt presented a neural network named Perceptron as a model of the brain (Rosenblatt 1958), and Selfridge presents the Pandemonium (Selfridge 1958), as a model of decision process for pattern recognition. In 1961 Steinbuch proposed matrix architecture (Steinbuch 1961) for studying neural learning. Studies of learning and pattern recognition abilities of perceptrons carried out during 1960s represent the enthusiastic period of neural nets research. However, in 1969 Minsky and Papert suggested that Perceptrons aren't a very promising direction (Minsky and Papert 1969). Afterwards, neural networks research teams received very limited or no funding. Even so, people continued to study the neural networks parading either individually or in isolated groups. In 1980 ANW (Adaptive Neural netWorks) group was formed at the University of Massachusetts/Amherst, with the idea of studying adaptive systems through the neural networks paradigm. In 1986, another group, the very well known PDP (Parallel Distributed Processing) group published the PDP volume (Rumelhart, McLelland and the PDP Group, 1986) studying parallel distributed processing by utilizing neural networks paradigm. Since then the Connectionism (neural networks) goes through renaissance period. While the PDP group was focused on advice learning paradigm, the ANW group was focused on reinforcement learning paradigm (Mendel and Mclaren 1970, Widrow et al. 1973, Barto 1997). An important achievement of the 1981 ANW group (Klopf, Spinelli, Arbib, Barto, Sutton, Anderson, Selker, Bozinovski, Porterfield) was the solution of learning with delayed rewards (reinforcements) problem using neural network (Bozinovski 1981, 1982, Barto et al. 1983). It is part of the challenge of assignment of credit problem that was considered both by the ANW and the PDP group: how to backpropagate the evaluation from the environment once action is performed by an agent. The PDP group solved the problem by error (= teacher advice – learner action) backpropagation, while the ANW group solved the problem by reinforcement (evaluation of learner's action) backpropagation. The reinforcement backpropagation is known in psychology literature (as secondary reinforcement mechanism as proposed in psychology literature (Keller and Schoenfeld 1950). First ideas on Consequence Driven Systems theory (crossbar adaptive array neural network, crossbar learning algorithm, feelings and emotions in neural networks, learning in dungeons-and-dragons environments) were proposed inside the ANW group in 1980. The author of this paper had a pleasure of personally meeting some members of the ANW group, while her taking courses at UMass during 1995.

2. BEHAVIORAL ENVIRONMENT AND LEARNING

Behavioral environment

The basic assumption is that an **agent** (a robot, a bacterium, a mouse, …) **acts in an environment**. Over time, an abstract impartial observer outside both the agent and its environment observes a behavior of the agent with respect to the considered environment. We define behavioral environment as an environment in which a behavior of an agent is expressed. We can also assign the environment to a certain problem space relevant for some considered task. Hence, a class of AI problems involving behavior in a problem space can also be considered within this framework

Learning

Learning is a process that represents itself on at least two levels:

1) In the learning agent’s memory a portion of *knowledge is gained* which contributes toward *building an expectancy map*, from which the agent generates a **behavioral** *policy with respect to that environment***.**
2) An observer observes that the entropy of the agent’s behavior is decreased. The behavior shows a pattern, from possibly totally random to possibly totally deterministic behavioral trajectory.

A learning agent is able to express a *learning behavior* in behavioral environments.

There are many **definitions of learning** and we will define learning as a relevant knowledge gathering using learning agent sensors in order to build and maintain a model of reality. The scope of reality is left underfund, but a limited scope is the environment the agent is interfacing.

3. A TAXONOMY OF LEARNING SYSTEMS

Learning system is an agent-environment interaction system that shows evidence of learning, either by knowledge update in the knowledge base of the agent or in decreasing entropy of agent’s behavior toward some goal oriented (purposive) behavior. A learning system consists of an agent (learner), an environment and the interaction between them. Among other taxonomies of agents (Franklin and Graesser 1996) and learning systems (e.g. Russel and Norvig 2010), here we present the taxonomy proposed by the Consequence Driven Systems theory (Figure 1).

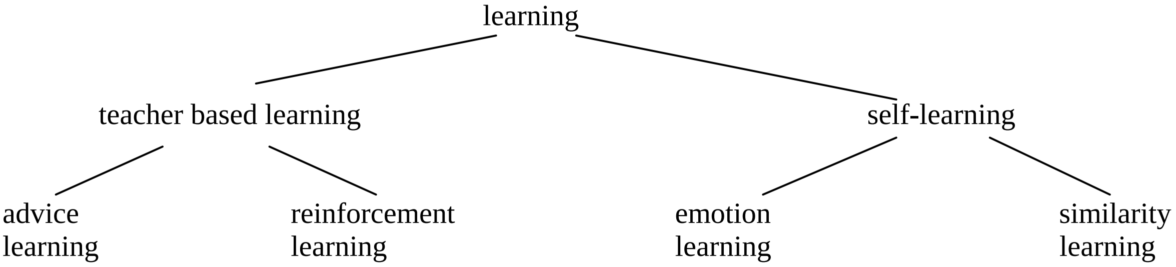

Figure 1. General taxonomy of learning systems

This taxonomy differs from the mainstream taxonomy of learning agents (supervised learning, unsupervised learning, reinforcement learning) since it explicitly considers the class of self learning systems, including emotion learning and similarity learning systems.

According to the taxonomy shown in Figure 1, the learning systems are divided into two main categories. The first category are teacher-based learning (supervised learning) systems in which the environment gives explicit evaluation (reinforcement) of the agent's behavior or/and explicit advice about future behavior to the agent. The second category contains self-learning systems, the systems that learn without any advice and without any reinforcement from the behavioral environment. According to CDS theory, such type of learning is only possible if the system has *genetically built* mechanisms that will allow self-learning. Therefore, in addition to the behavioral environment, this theory introduces the genetic environment in order to study the self-learning systems. Once the genetic environment is introduced, a self learning system uses two genetically built evaluation functions: 1) one such function is emotion evaluation 2) the other such function is similarity evaluation. Based on that, the category of self-learning systems contains emotion based as well as similarity based learning systems.

The self-learning systems based on similarity evaluation are able to build clusters (or classes) of objects and in that way to develop cognitive concepts about the environment. The similarity based self-learning is evident in some tasks of cluster analysis in pattern recognition problems. This paper does not elaborate further on similarity based self-learning. Here emphasize will be more on emotion based self learning systems.

4. THE AGENT-ENVIRONMENT INTERFACE

This part presents some formalism about agent environment interaction and then discusses some learning paradigms. It is assumed that a learning agent receives a set ***S*** of signals which represent the environment, and acts toward the environment with a set ***Y*** of signals which is considered as an agent's behavior toward the environment (Figure 2).

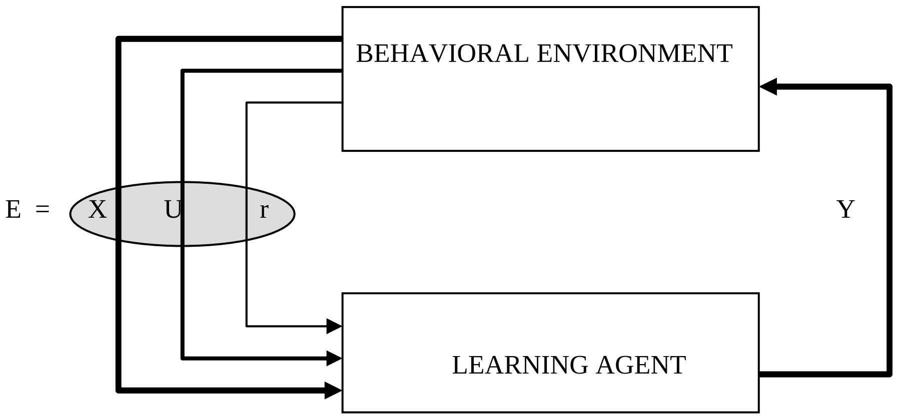

Figure 2. The agent-environment interface

The behavior ***Y*** could be a sequence of actions, or just a single action. Situation in moment **t** is a model of the environment as presented by the scan of the sensors in moment t. Situation is concept built by the cognitive capacities of an agent. Action is integrative output of agents actuators executed in moment t. Learning can be viewed as building a policy (situation, action). It is not just a model of the environment, but a behavioral policy how to behave in the environment.

The set ***E*** is denoted as *generalized situation* and represents the *behavioral environment.* It consists of three subsets,

$$E = \{X, U, r\} \tag{1}$$

where

X is a *behavioral context*, the set of signals E-$\{r, U\}$; usually considered as "neutral" signals; ***X*** will be referred to as *situation.* In the situation **X** the agents generates a behavior **Y**. Both **X** and **Y** are sometimes written as vectors rather then as sets.

r is *behavior reinforcement* generated by the environment. It is a distinguishable signal in ***E*** which is presented to the learner to show how good the learner's performance is in the considered environment; it is a *scalar* variable (e.g. grade, salary, food) which the agent will tend to *optimize* in the process of learning. This means that *reinforcement is agent's behavior evaluation, evaluated by the environment and presented to the agent.*

U is a *behavior advice.* It is a distinguishable signal, or set of signals, generated by the environment to provide *advised behavior in a particular situation* (policy) for the learning agent.

The set

$$LI = \{E, Y\} \tag{2}$$

is named *agent-environment interface* between the environment and a learning agent. In some cases it is convenient to consider LI as an ordered n-tuple, rather then as a set, to emphasize the *sequence of events* appearing in the learning process.

5. A TAXONOMY OF LEARNING PARADIGMS

The following definitions are used in description of the taxonomy of learning agents and paradigms.

Learning paradigm is a sequence of interaction steps that occur between the agent and the environment in a learning trial.

Learning trial is a unit of a learning experiment.

Learning experiment is a sequence of steps in which the learning paradigm is iterated. If the iteration converged to a point that the learner does not change the behavior any more in respect to the inputs given in a learning trial, it is said that the learning process converged. The learner updated its knowledge base so that from that point on, it will generate future behavior according to the learned knowledge.

In a simulation of the learning process, a learning trial is an iteration of the computer program that includes memory update due to learning.

In the sequel the paper uses the following notation for describing learning paradigms:

A denotes agent,

E denotes environment,

x, y, u, r are current situation, action, advice and reinforcement respectively,

x', u', are next situation and advice respectively.

Context-free paradigms

A context-free interface does not consider the (neutral) situation x. Here we consider two such paradigms, forced training and reinforcement based training.

LI 1*: **Context-free advice learning paradigm***: u-**A**-y-**E**-u'-

In this paradigm the environment in the role of trainer just gives the advice u and the learner learns to follow that advice with its behavior y. After several learning trials, y converges toward u.

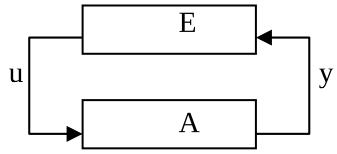

Figure 3. Context-free advice learning: u-**A**-y-**E**-u'-

There is a teacher that gives advice as to what behavior should be performed. The learner can repeat the actions advised by the teacher. Example of such learning is repeating a phrase of words that teacher requires learner to learn. Another example is learning by imitation.

LI 2: ***Context free reinforcement learning paradigm:*** **A**-y-**E**-r-**A-**

In this paradigm the learning agent tries actions y in order to optimize the reinforcement signal r.

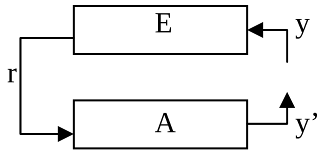

Figure 4. *Context free reinforcement learning*: **A**-y-**E**-r-**A-**.

This learning paradigm is used in learning automata and function optimization. This is a primitive type of a consequence learning paradigm.

Context-sensitive paradigms

In context sensitive paradigms, the (neutral) situation in which a situation is perceived, is taken into consideration. Association between the situation and the teaching signal (either advice or reinforcement) is being built.

LI 3: ***Association learning paradigm*:** (x, u)-**A**-y-**E**-(x', u')

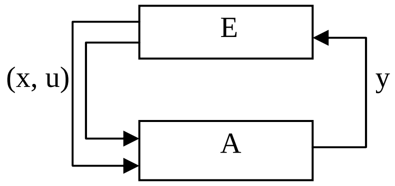

Figure 5. *Association learning*: (x ,u)-**A**-y-**E**-(x',u')

In this paradigm, both the situation *x* and the advice *u* appear simultaneously, as a pair. The teacher advises the learner that given the situation x, instead of association ($x \rightarrow y$) exhibited by the student, a desired association ($x \rightarrow u$) should be established. After the learning is completed the student exhibits the learned association ($x \rightarrow u$). This type of learning paradigm is used in content addressable memories (Spinelli 1970) and in pattern recognition systems.

LI 4: ***Classical conditioning paradigm***: (*x ; u)-**A**-y-**E**-(x'; u')-*.

In this paradigm it is important to emphasize that between *x* and *u* there is a time difference, denoted as (x ; u).

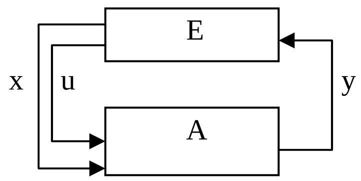

Figure 6. *Classical conditioning*: (*x ; u)-A-y-E-(x';u')-*

This is the well known *classical conditioning* paradigm (Pavlov 1927). In a context x (conditioning stimulus, CS), an advice u (unconditioned stimulus, US) is applied which produces a behavior y. After several trials, an association between x and y is learned, so when situation x appears the behavior y is produced by the learner.

Consequence learning paradigms

In the following series of paradigms the leaning appears as a consequence of previous behaviors.

LI 5**:** ***Reinforcement learning paradigm***: *x-**A**-y-**E**-r, x'*

In this paradigm, in a situation x the agent performs action y related to situation x, and the environment returns a distinguishable signal r that is interpreted by the agent as reinforcement. This is the classical context-dependent reinforcement learning paradigm: the agent learns to associate the appropriate action y to the situation signal *x,* in order to optimize the reinforcement *r.*

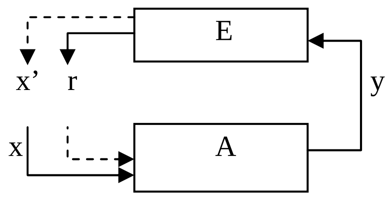

Figure 7. *Reinforcement learning paradigm*: *-A-y-E-(r, x')-*

This paradigm is a model of the operant (instrumental) conditioning paradigm in animal learning theory (Skinner 1938): Given a lever as signal (instrument) x, the animal performs action y (push the lever) after which food appears as reinforcement r. It is a consequence-learning paradigm without an advice-giving teacher. Some authors in modern human psychology indeed understand the *operant conditioning as learning based on consequences* (Baron 1999). Note that in this paradigm a teacher is present, just instead of giving an advice, it gives an evaluation of behavior, a reinforcement, in this case a reward. Note also that in the initial action of the learner (e.g. animal) is a cognitive process such as curiosity or/and motivation.

LI 6: **Error-correction learning paradigm.** *x-**A**-y-**E**-u-x'-*

The learner receives the situation x, produces an action y, and receives from the environment an *advised action u as consequence of the produced action y in situation x.*

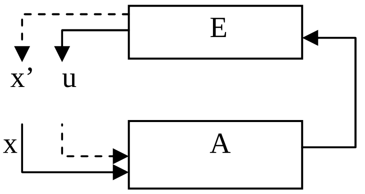

Figure 8. Error-correction learning paradigm: *x-**A**-y-**E**-(u-x')*

This paradigm is used in some schemes of error-correction learning, such error-backpropagation learning (Rumelahrt et al. 1986). The environment (teacher) gives advice *u* related to the action *y* received by the learning agent.

LI 7: ***Reinforcement with selected advice learning paradigm*:** - *x-**A**-y-**E**-(r, [U])-x'-.*

In this paradigm the advice u is given by the environment only if needed. Evaluation r is used to evaluate agent's action y on situation x. Hence, evaluation r is applied in every teaching trial, but advice is given only if needed, not in each step.

To decide about the need of introducing a *teaching trial*, the paradigm assumes *examination trial*. If in an examination trial the learner does not respond properly, it first receives a negative reinforcement r, (for example the word "No!"), and after that, advice as to what to do.

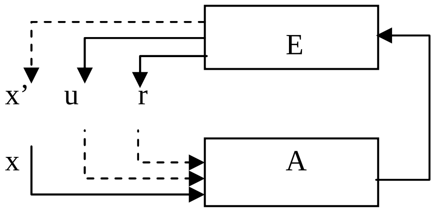

Figure 9. Reinforcement with selected advice learning paradigm: - *x-**A**-y-**E**-(r,[U])-x'*

The learner tends to adjust its x-*y* mapping toward the requested x-*u* mapping, and also to minimize the amount of negative reinforcement received. This paradigm has been used in teaching for pattern recognition and lessons learning tasks.

LI 8: ***Delayed reinforcement learning paradigm:*** x_1-**A**-y_1-**E**-x_2-**A**-y_2-**E**-[r_1]-x_3-

In this paradigm, the learner receives evaluation (reinforcement) on its action y performed several situations before, not for the actions performed in the previous situation. For example, reinforcement r may appear two steps after action y has been executed, as shown in Figure 10.

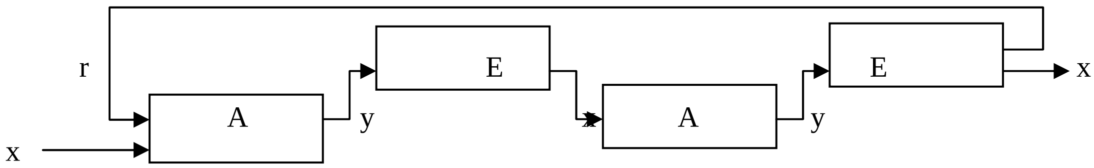

Figure 10. Delayed reinforcement learning paradigm - x_1-**A**-y_1-**E**-x_2-**A**-y_2-**E**-[r_1]-x_3-- '

This is a *delayed reinforcement learning paradigm.* In some cases, like in the game of chess, the reinforcement is received only at the end of the game.

LI9: ***Delayed advice learning paradigm**:* x_1-**A**-y_1-**E**-x_2-**A**-y_2-**E**-$[u_1]$-x_3-

This is a *delayed advice learning paradigm,* and *is s*imilar to delayed reinforcement learning paradigm. Here the advice u on (x, y) is received several steps after learning step (x, y) has occurred ("in that situation x you should have done y").

Self-learning paradigms

In a self learning interface there is neither reinforcement r nor advice u received from a teacher. The environment only gives (neutral) situation x.

LI 10**: *Emotion learning paradigm:*** *x*-***A***-*y*-***E***-*x'*-.

In this learning interface the environment presents only (neutral) situations. Neither advice nor reinforcement is given, not even a delayed one.

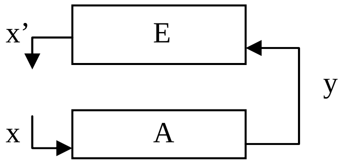

Figure 11. *Situation-value (emotion) learning paradigm*: x-**A**-y-**E**-x'-.

This is the case of a *self-learning paradigm*. The learning agent should develop its internal evaluation mechanism in order to exhibit a learning behavior. Such a mechanism is the emotion (feelings) mechanism. The situation is considered as a representation of the state of the environment, and the concept of *state evaluation* was introduced (Bozinovski, 1982) in theory of learning agents. The concept of state evaluation is also used in Dynamic Programming (Bellman, 1957). The concepts of Dynamic Programming such as state value and action value were used (Watkins 1989) to develop a modern theory of learning from delayed rewards. Currently, emotion as state evaluation is part of appraisal theories of emotion (Brave and Nass, 2003). Other researchers (Gadanho, 2003) also relate emotions with the learning process.

Emotion is considered important part of the educational process as well. The Bloom's taxonomy (Bloom et al, 1956) of educational objectives in the new revisions (Marzano and Kendall, 2007) addresses the self-part of systems including emotion.

6. SELF-LEARNING AGENTS WITH EMOTION SELF-EVALUATION

The taxonomy of learning interfaces presented above assumes existence of learning agents capable of learning in a particular learning interface. Figure 12 shows three types of agents and their interfaces to the environment.

Most abundant interface for an agent is interface shown in Figure 12.a. The environment supplies a situation X, an evaluation r of the previous action (behavior), and also an advice U for a proper behavior in the considered situation. The agent which learns in such environment is named class T (tutorial) agent; it learns with guidance of a teacher (an oracle that knows all the answers). Consequently such oracle guided agent need not to develop sophisticated learning physiology, i.e. mechanisms that will allow learning. The second agent, shown in Figure 12.b. receives from the environment only evaluation of its previous behavior. Therefore, such an agent should at least remember its previous behaviors. It needs a physiology for recognition of an external reinforcement and association of that reinforcement with its own behavior. This agent is of R (reinforcement) type.

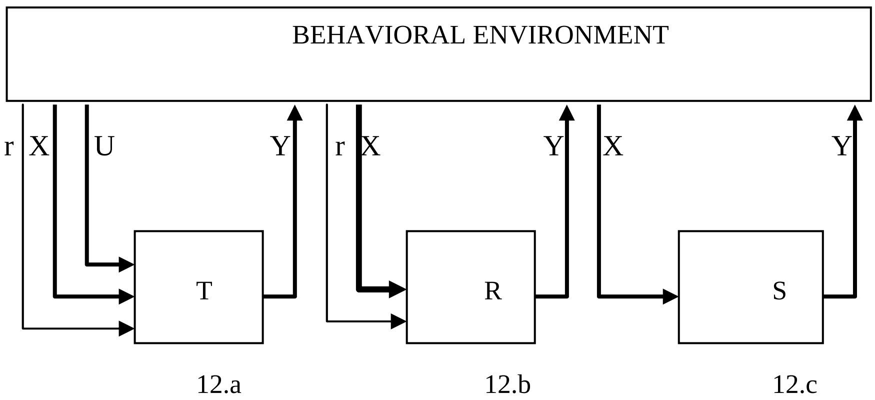

Figure 12 Classes of context-dependent consequence learning agents

The most complex learning physiology needs the agent of class S (self-learning) shown in Figure 12.c. This type of agent receives only situation as its input. It needs a physiology to recognize that the received situation is a consequence of its previous behavior. It also needs a physiology to evaluate that situation, with a mechanism of emotion. It might be a mechanism for computing desirability (state evaluation) of being in that situation. Such a mechanism can be genetically built. Hence, in addition to behavior environment, a self-learning agent needs a relation to a genetic environment that will establish initial concepts of danger, fear, pleasure, pain, cold, and similar basic surviving emotions where from, by *emotion backpropagation* (secondary reinforcement) mechanism, the agent will develop learning behaviors. Computation of an emotion and use of that emotion as evaluation mechanism is essential in building a self learning agent.

7. LEARNING AGENTS MORPHOGENESIS BY AXON REWIRING

Within the theory of Consequence Driven Systems one can study the morphogenesis of learning agents from class T agents (capable of learning only with an advisor), to class R agents (capable of learning with as little as reinforcement from the environment) and finally to class S agents (capable of self-learning, with no advice and no reinforcement from the environment).

The approach to morphogenesis is based on two principles: The first principle is that learning agents have neural architectures. The second principle is that those neural architectures have interface to both the behavioral environment and to the genetic environment. Such an agent is named *neurogenetic agent (NGA).* The steps of morphogenesis are shown in Figure 13.

The T type agent in Figure 13 is the basic neurogenetic agents. It can be developed from its genotype as it is discussed elsewhere (Ackovska et. all 2008). The type T agent can learn only in an environment that provides both advice and evaluation (reinforcement) about agent's behavior.

Figure 13 shows how a T type agent can undergo morphogenesis and transform into a self learning agent. The basic principle of morphogenesis is axon rewiring.

To obtain a reinforcement-learning (R) agent from a tutoring-learning (T) agent, axon collateral from output axon Y connects to the teaching (U) input of the agent. Hence, after rewiring we obtain by the architectural design that U(t) = Y(t-1). That means the advice for future behavior is: in situation X perform the same action as the action Y(t-1) performed in the last step (t-1) in that situation. Such advice will be stored in memory only if reinforcement signal r(t) is a rewarding signal. If it is neutral or a punishment signal, the situation-action pair will not be remembered. One can observe that a simple feedback loop from the action output to the advice input is all that is needed for a class *T* (advice-learning) agent to be transformed into a class *R* (reinforcement-learning) agent.

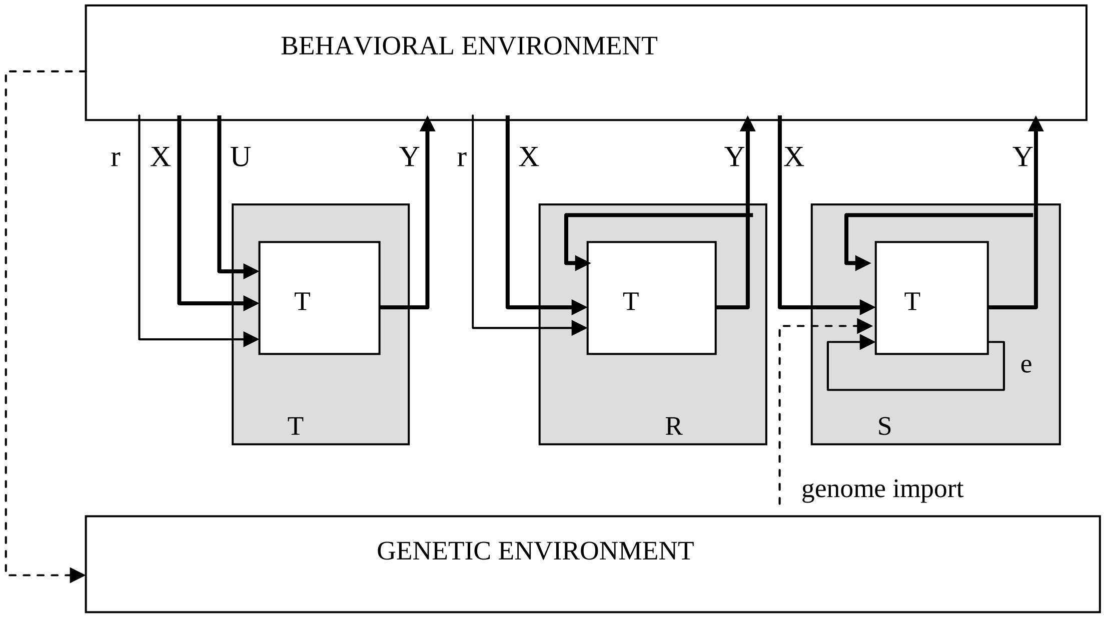

Figure 13. Morphogenesis of learning agents. From T class into R class and finally to S class agent

To obtain a self-learning (S) agent from a reinforcement-learning (R) agent, it needs a mechanism for emotion computation. Therefore, the S system uses crossbar-adaptive array architecture (Bozinovski 1982, 1982) that in crossbar fashion computes both its actions and its emotions from the same memory array. Given situation X(t) the memory computes *e*(X(t)) which is desirability of being in situation X. Hence, the output of the CAA consists of both action *y*(X(t)) and emotion *e*(X(t)) due to received situation X(t). The action is fed back to the teaching input, as needed to obtain an R learning system. A second feedback is from the emotional output e(t) to the evaluation (reinforcement) input r(t). One can observe the obtained S learning agent uses its emotion as evaluation of a previously taken action. The obtained system is capable of learning with no teacher at all: using its emotional evaluation of the received situation X, it will chose use or not chose its previous action as advice for future action policy in situation X.

The input from the genetic environment (genome) is needed to establish the initial inherited emotions of the agent. For example the feeling “it is cold” should be genetically related to undesirable emotion, otherwise the agent will not survive in the environment. Genetic environment is in relation to behavioral environment and provides initial information to a newly born agent about surviving in that environment. According to this theory both initial genetic emotion and emotion backpropagation mechanism are needed for self-learning to occur.

8. CONCLUSION

This paper shows further elaborations on taxonomy of learning agents based on Consequence Driven Systems theory. There are various types of learning paradigms, and this paper elaborates on their taxonomy.

Advice-based learning is a subset of learning paradigms, which assumes an advice provider (oracle, teacher, book, etc) that tells the learning agent what to do in a particular situation. Usually a teacher has a set of lessons (patterns, skills, lectures), and those lessons are presented to a student.

Reinforcement-learning paradigm needs no advice-giving environment, and relies only on the evaluation-giving environment. It needs a memory to store its previous action performed in a previous situation. If evaluation from the behavioral environment received as consequence of the mentioned situation-action pair is rewarding, the association situation-action will be formed as a part of learned behavioral policy in the considered behavioral environment.

The most advanced learning agents are the self-learning agents that are capable of learning with no teacher (neither advice-giving nor reinforcement-giving). Emotional self-evaluation of encountered situations, as well as emotion backpropagation are needed features in the physiology of a self-learning agent. The initial emotions toward some states of the behavioral environment are given from genetic environment. The genetic and the behavioral environment should be in accurate correspondence about the conditions in the behavioral environment in order an agent to be allowed existence and learning in the considered behavioral environment

Using axon rewiring mechanism the CDS theory shows morphogenesis of learning agents. It shows how an agent can evolve from an advice-learning agent, to reinforcement-learning agent, and also to self-learning agent.

REFERENCES

Ackovska N., Bozinovski S., Bozinovska L. (2008) Evolving an emotion-based neural control architecture. Proc. IEEE SoutheastCon, Richmond, VA, p. 404-408

Baron R. (1999) Essentials of Psychology. Allen and Bacon

Barto A. (1997). Reinforcement learning. In O. Omidvar, D. Elliot (eds.) *Neural Systems for Control*, Academic Press, p. 7-30

Bellman R. (1957) Dynamic Programming. Princeton University Press

Bloom B., Engelhart M., Furst E., Hill W., Krathwohl D. (eds.) (1956) Taxonomy of Educational Objectives. The Classification of Educational goals. Handbook I. Cognitive Domain. David McKay

Bozinovski S. (1981) Inverted pendulum learning control ANW Memo, Adaptive Networks Group, Computer and Information Science (COINS) Department, University of Massachusetts at Amherst, December 10, 1981

Bozinovski S. (1982) "A self-learning system using secondary reinforcement" in R.Trappl (Ed.) *Cybernetics and Systems Research*, p. 397-402, North Holland

Bozinovski S. (1985) Adaptation and training: A viewpoint. Journal Automatika 26 (3-4): 137-144

Bozinovski S. (1995) Consequence Driven Systems, Gocmar Press, Athol, MA

Bozinovski S., Bozinovska L. (2001) Self-learning agents: A connectionist theory of emotion based on crossbar value judgment. Journal of Cybernetics and Systems (32):637-669

Brave S., Nass C. (2003) Emotion in Human-Computer Interaction In J. Jacko, A. Sears (eds.) handbook of Human Computer Interaction. Lawrence Erlbaum Associates, p. 82-94

Gadanho S. (2003) Learning behavior-selection by emotions and cognition in a multi-goal robot task. Journal of Machine Learning Research 4: 385-412

Franklin S., Graesser A. (1996) Is it an agent or just a program: A taxonomy of autonomous agents. Proc. ECAI Workshop on Intelligent Agents III: Agent Theories, Architectures, and Languages. p. 21-35, Springer Verlag

Keller F., Shoenfeld W. (1950) Principles of Psychology. Appleton-Century-Kroffts

Marzano R., Kendall J. (2007) The New Taxonomy of Educational Objectives, Corwin Press

McCulloch W., Pitts W. (1943) A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics 5: 115-133

Mendel J., McLaren R. (1970) Reinforcement learning control and pattern recognition systemds In J. Mendel, K-S. Fu (eds.) Adaptive, Learning, and Pattern Recognition Systems: Theory and Applications. Academic Press, p. 287-318,

Minsky M., Papert S. (1969) Perceptrons. MIT Press

Pavlov, I. (1927) Conditioned Reflexes, Oxford University Press

Rosenblatt F (1958) The perceptron: a probabilistic model for information storage and organization in the brain. Psychological Review 65: 386-408

Rumelhart D., McLelland J., and the PDP Group (1968) Parallel Distributed Processing. Explorations in the Microstructure of Cognition. MIT Press, 1986

Russel S., Norvig P. Artificial Intelligence: A Modern Approach. Prentice Hall, 2010

Selfridge O. (1958) Pandemonium; A paradigm of learning. Proc. Symposium on Mechanization of Thought Process. HMSO, London, UK, p. 513-526

Skinner B. (1938) The Behavior of Organisms: An Experimental Analysis. The Skinner Foundation

Spinelli N. (1970) OCCAM: A computer model for a content addressable memory in the central nervous system. In K. Pribram, D. Broadbent (Eds.) Biology of Memory, Academic Press, p. 293-306

Steinbuch W. (1961) Die Lernmatrix Kybernetik 1: 36-46

Watkins C. (1989) Learning from delayed rewards, PhD. Thesis, Kings College, London

Widrow B., Gupta N., Maitra S. (1973) Punish/Reward: Learning with a critic in adaptive threshold systems. IEEE Transactions on Systems, Man, and Cybernetics 3(5): 455-465