Mental development and representation building through motivated learning

Janusz STARZYK

Pawel RAIF

Ah-hwee TAN
Singapore Management University, ahtan@smu.edu.sg

Citation

STARZYK, Janusz; RAIF, Pawel; and TAN, Ah-hwee. Mental development and representation building through motivated learning. (2010). *Proceedings of 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, July 18-23*. 1-8.
Available at: https://ink.library.smu.edu.sg/sis_research/6773

Mental Development and Representation Building through Motivated Learning

Janusz A. Starzyk, *Senior Member, IEEE*, Pawel Raif, *Student Member, IEEE*
and Ah-Hwee Tan, *Senior Member, IEEE*

Abstract—Motivated learning is a new machine learning approach that extends reinforcement learning idea to dynamically changing, and highly structured environments. In this approach a machine is capable of defining its own objectives and learns to satisfy them though an internal reward system. The machine is forced to explore the environment in response to externally applied negative (pain) signals that it must minimize. In doing so, it discovers relationships between objects observed through its sensory inputs and actions it performs on the observed objects. Observed concepts are not predefined but are emerging as a result of successful operations. For the optimum development of concepts and related skills, the machine operates in the protective environment that gradually increases its complexity. Simulation illustrates the advantage of this gradual increase in environment complexity for machine development. Comparison to reinforcement learning indicates weakness of the later method in learning proper behavior, even in such protective environments with gradually increasing complexity. The method shows how mental development stimulates learning of new concepts and at the same time benefits from this learning. Thus the method addresses a well know problem of merging connectionist (bottom-up) and symbolic (top down) approaches for intelligent autonomous machine operation in developmental robotics.

I. INTRODUCTION

A RTIFICIAL neural networks were used in many learning and classification tasks and can produce a robust, performance to recognize objects or sequences of sensory input data useful for building concepts as well as semantic and episodic memories. It is obvious that neural structures in the brain perform all advanced level processing of information including concept building, learning new skills, developing motivations to act, setting objectives, planning and thinking. They respond to external stimuli to the same extent as to their internal motivating signals. Yet, these capabilities are currently out of reach of artificial neural networks. There is a significant gap between what artificial neural networks (connectionist networks) are capable of doing in a complex system and the growing need to manipulate information extracted through these networks. While the information that enters a neural network input often represents low level sensory data (like pixel intensity), its output may correspond to categories with robust invariant properties. Although, these categories are useful for symbolic manipulation, there is no natural extension of neural network learning and organization that would yield such manipulation. Our work tries to address this problem by offering an approach that combines concept development with its use in the motivated learning system.

The reinforcement learning (RL) mechanism [1] uses a reward signals to train the machine act in a desired way. The machine is organized in such a way that it maximizes the expected reward decided by and received from the external environment. By selecting and controlling this reward a designer has a complete control over machine behavior and even an optimum reinforcement learning machine can be build (at least in theory) [2]. In addition, RL suffers from the curse of dimensionality and computational cost to learn increases significantly with the environmental complexity [3]. Hierarchical reinforcement learning with subgoal creation was introduced to help alleviate these problems with some success [4] [5]. The hierarchical reinforcement learning exploits the environment and organizes the agent's tasks to improve policy learning. Using subgoals to build a hierarchy of subsequent goals improves the efficiency of RL. Bakker and Schmidhuber [1] proposed a method for hierarchical reinforcement learning based on subgoal discovery and subpolicy specialization. Their reinforcement learning algorithm created both useful subgoals and the specialized subtask solvers. This approach may yield a complex hierarchy of subgoals, although a large number of parameters and lack of strict convergence guarantees are weak points of their approach.

To stimulate machine's autonomous development intrinsic motivation mechanism was proposed based on curiosity-driven exploration, novelty and surprise. This phenomenon known in psychology and neuroscience [6] was used by Schmidhuber to develop artificial curiosity in robots for exploratory behavior in unknown environments [7].

In a similar effort based on the curiosity principle, Oudeyer [8] proposed an intelligent adaptive curiosity (IAC) system. In this system a machine needs to identify unpredictable response or sequence of responses to learn either the

environment properties or its own actions. Artificial curiosity is the motivating force behind learning and the associated mechanism helps to establish new concepts and develop new skills. Curiosity based observations provide sensory and motor categories that speed up the reinforcement learning process. The learning process was compared to that of young children who learn by playing, and change their focus in the game to new observations and skills. Although this learning is mostly guided by child's own interests, it may be used later to advance its specialized knowledge.

Neural network implementation of intrinsic motivation principle in machine learning would require comparing prediction of the sensory input to the observed input and a mechanism to minimize the prediction error (Kaplan and Oudeyer [9]). Such approach seems to have a strong support in the observed dopamine regulating structures in the brain. One hypothesis is that tonic dopamine acts as a signal of expected prediction of error decrease and the second one considers cortical microcircuits that acting as prediction systems. Human based experiments were proposed to verify Kaplan and Oudeyer's hypotheses.

Other authors focused on additional features of curiosity based learning, like liking intrinsic curiosity based reward with extrinsic goal related reward (Roa et al. [10]) pointing out problems in formulation of tasks from observed temporal experiences. In [11] authors acknowledge inherent value system as initial drive for autonomous development of the agent. They introduce computational model of such initial value system. Their proposed system integrates three types of signals: punishment (as an example of aversive stimuli), reward and novelty. As they state the punishment value has the highest weights in their system. It resembles our concept of internal pain signals that produce rapid change in agent's behavior.

Vernon and co-authors [12], present a comprehensive analysis of research in developmental systems pointing towards requirements that must be satisfied by such systems like: competing and cooperating subsystems, adaptable and self-organizing architecture (both parameters and structure), ability to predict and verify observations and environment response, and the embodiment of mind to carry interactions with the environment necessary for developmental cognitive systems. Such systems develop incrementally and its representation of environment is gradually enriched together with its emerging abilities to act.

By using intrinsic motivations, a machine can explore the environment and learn complex skills. Such skill may later be found useful to improve its performance in the environment [13]. Intrinsic motivation as used in curiosity based learning is similar to exploration in reinforcement learning (RL). In RL a machine does not always respond in an optimum way but occasionally tries a random search in state-action space. Intrinsic motivations favor actions that minimize the prediction error. Thus if a machine focuses on one type of action, it may improve its operation using intrinsic motivations in a similar way as random exploration improves

performance in reinforcement learning. However, if no task is specified, curiosity based learning provides general knowledge about the environment, which may or may not be useful from the point of view of the machines objectives. Once a machine needs to specialize, such exploration slows down its progress and overloads its memory with unnecessary observations. Thus there is a need for internal, goal orientated motivations in machine learning.

In this work we focus on the role of the motivated learning in building cognitive representations of objects and concepts that machine finds useful for its mental development. We will illustrate how this development takes place enriching machine's ability to act and comprehend the environment. We relate this development to changes in the environment that stimulate machine to learn and show that systems that do not use motivated learning may miss the opportunity to learn advanced concepts and, as a result, will not be able to compete. Finally, we show dependence between environment conditions and learning, indicating that learning can be accelerated by properly controlled changes in the environment.

II. MOTIVATED LEARNING

A fundamental problem of embodied systems, that use intrinsic motivations in its development, is that they do not have a natural mechanism that links intrinsic motivations to externally set goals. Attempts to combine the knowledge learned by the machine through curiosity based explorations to external goals are limited to designer's effort rewarding the machine's actions whenever he/she seems appropriate [10]. Although such efforts are useful and improve the learning efficiency, they cannot be a substitute for an internal drive to learn for a purpose. Intrinsic, curiosity based motivations are fully internal, but are disconnected from the externally set goals. Externally rewarded behaviors are goal oriented but provide limited motivations to explore and learn new abstract objectives and concepts.

While curiosity based learning is useful for the early stage of robot development, it may slow down learning of specific skills that may be required for complex tasks. Subgoals discovered in hierarchical reinforcement learning (HRL), are obtained by clustering input data [1] to arrive at desired and useful results. Discovered subgoals are later used as stages needed to accomplish a complex goal. Such approach assumes that the environment is stationary, so the subgoals in HRL are specialized partitions of the input space to facilitate learning of the value function. Although automatic subgoal creation and selection facilitate operation and improves the learning speed in RL, it limits learning to prespecified objectives for which rewards are externally given. This constrains machine's development.

Motivated learning (ML) mechanism proposed in [14] was designed to provide motivations to the learning machine that combine its externally driven goals with internal goals emerged from the developmental process controlled internally

by the machine. Motivated learning first learns dependencies between objects in the environment and the externally set objectives (controlled by the external rewards), and subsequently, uses these observations to set internal goals. The machine responds to specific goals, trying to find solution to the problem set, so it explores the environment with a specific objective. Motivated learning uses negative reward systems as its major reinforcement. Negative signals (also known as pain signals) are received from the environment and need to be minimized (synonym of reward). If the negative signals increase instead of being reduced, this corresponds to a negative reward and machine learns not to perform action that resulted in such increase.

Several pain signals compete for machine's attention at any given time, and the machine concentrates on the strongest signal learning how to minimize this signal. If all the pain signals are below specified threshold, the machine switches to curiosity based exploration. Such exploration helps to learn basic dependencies in the environment and some of them may be useful for its operation. However, the main effort in motivated learning is concentrated on learning desired behavior that addresses specific needs. Thus learning is focused, and both attention and learning is switched automatically by competing pain signals.

This ML mechanism that triggers desired action based on the observed input is the foundation for building representations of the external world. In fact objects are defined first of all as resources needed to perform desired tasks. The same mechanism helps to build stable representations associating similar looking objects with desired actions. This defines object equivalences, yielding their invariant representations. New concepts created by ML machine lead to new motivations. The main strength of ML is that it produces value systems related to many abstract concepts in the environment and relates them to its objectives, without receiving an explicit reward for this learning.

A. Characterization of Motivated Learning

Motivated learning may be defined for both symbolic and connectionist approaches. Thus it is a good method to develop concepts and values, to learn observations and action starting from raw sensory input signals and delivering simple motor control. Its pain based structure was illustrated in [14] using simple neural circuits. Such circuits can evaluate changes of the pain signal and create abstract goals based on the observed useful actions. This introduces abstract pains and motivations to accomplish machine created abstract goals.

Definition:
Motivated learning (ML) is learning in an embodied agent based on a self-organizing system of emerging goal oriented internal motivations.
• ML creates abstract motivations and chooses goals based on the primitive pain signals.
• It receives internal rewards for satisfying its goals (both primitive and abstract).

• Environment changes stimulate agent's development.

A ML machine is in a continuous process of building new motivations and responding to established ones. Competing signals, that represent abstract pains and attention-switching, direct the machine to choose a goal to act on and to follow this goal. These signals vary as the machine acts and the environment around it changes.

The ML mechanism triggers learning of intentional representations and establishes associations between sensory observations and motor actions. Thus the machine responds to the observed environment changes and to its own internally generated pain signals to chose proper action. This response is as much a result of top down deliberation and prediction of what will be the result of its action as well as its bottom-up perceptions, experiences and past history of interactions with the environments.

This learning supports cognitive functions, defining abstract categories if they are found useful for the machine's operation. Such useful object categories are learned better and faster than objects discovered through a curiosity based search. In ML mental development results from this process of representation building and goal creation, but at the same time, it is used to refine its representation building and learn new actions.

The external pain signals are predefined and connected to pain detection centers that trigger the learning mechanism. Thus, the machine's motivation comes from its response to external pain signals. Intelligence cannot develop without embodiment or interaction with the environment. Through embodiment, intelligent agents carry out motor actions and affect the environment. The response of the environment (including the pain signals) is registered through sensors implanted in the embodiment. At the same time the embodiment is a part of the environment that can be perceived, modeled and learned by the intelligent machine.

Although both reward and punishment signals can be used to stipulate learning, avoiding punishment may be sufficient for an agent's development (at least in simpler systems) and unlike reward maximization will lead to stable systems. Maximization of total reward leads to a classical maximization problem and may produce unstable systems (with infinite reward); while actions that reduce the most negative pain signal will be terminated once the signal is reduced below specified threshold. In addition, the pain reduction provides a natural mechanism to manage motivations and goal selection in multi-objective systems. Thus the motivated learning uses chiefly a pain reduction mechanism.

B. Pain-based Goal Creation

Primitive pain signals are externally defined and generated and the machine needs to learn how to minimize them. In sophisticated environments there are rules that govern relationships between various objects that affect the machine's perception and in particular its pain signals. By discovering these rules and learning how to use them to its advantage the machine develops complex knowledge about the environment.

It will do so through internal motivations that relate solutions of its goals to desire for creating conditions in the environment under which such solutions are possible. Thus, the machine learns how to actively change the environment to its own advantage, rather than just responding to an existing state of the environment.

A simple neural network organization shown in Fig. 1 can be used to detect the change in the received pain signal and to activate learning. Growing pain signals compete with other pain signals for machine's attention in the winner takes all circuit. In this figure sensory and motor neuron activities are symbolically represented by single neurons for simplicity, although, distributed representations of sensory objects or motor actions will be used to build bottom up representations and top-down motor control.
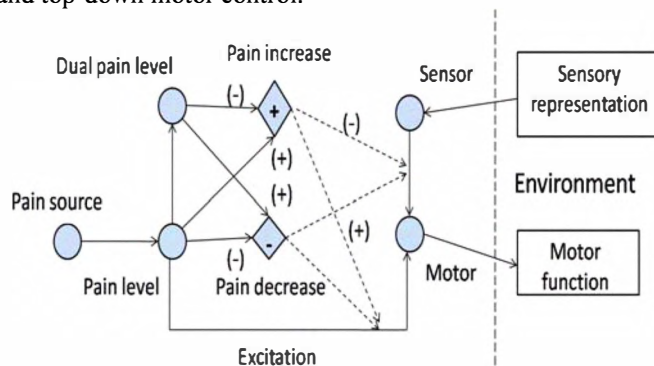

Fig. 1. Basic pain detection and learning unit.

The machine is directed by the winning pain signal to search for the proper action. If such a successful action is performed, the machine creates an abstract pain center that is activated if this action cannot be performed (e.g. due to lack of resources or impairment of the motor action). Thus the machine must evaluate, how likely it is that the needed resource will be hard to obtain or a desired action be impaired to assign an abstract pain value. The more often a specific resource was used to satisfy the machine's goals, the more important it is and the higher weight will be assigned to unavailability of this resource. Thus if the machine needs to use water to water plants (its major task for which it is rewarded), the lack of water is perceived as an abstract pain, and this pain intensifies with importance of water (frequent use) and difficulty to acquire it.

C. Representation Building

The winning pain signal forces the machine to explore its environment to find solution. Such solution can be found through exploration or observation of someone performing a desired task. In doing so, machine discovers relationships between objects observed through its sensory inputs and actions it performs. Observed concepts are not predefined but are emerging as a result of successful operations. Thus a concept of an object is related to useful and predictable properties this object may have with regard to the machine's objectives. In connectionist networks objects are recognized

mostly through self-organization of similar features and a feature invariance building through continuous observations. For the optimum development of concepts and related skills, the machine operates in the protective environment that gradually increases its complexity. Thus, the developmental process must be monitored and the learning environment structured in such a way as to facilitate the machine learning.

An important observation is that representation building which results from association of observed actions with the internal or external reward comes from motivation of the machine to act, while in turn motivations to act come from representation building. New representations may yield new motivations to protect or acquire desired resources, while new motivations force machine to discover new ways of solving its problems, leading to new representation building, new motor skills and new semantic and episodic memory.

D. Abstract Pains, Motivations, and Mental Development

Internal goals are created by the machine based on their relations to externally specified objectives. Thus, the machine learns causal relations between its internal goals and externally reinforced ones. By learning how to satisfy the external goals, the machine learns to anticipate an outcome of its action.

An abstract pain center uses a similar organization of pain detection and learning unit as shown in Fig. 1, to trigger abstract motivation. However, an abstract pain center is not stimulated from a physical pain sensor; instead the pain sensor only symbolizes an internal pain from not having sufficient resources to lower its primitive or abstract pain.

At any given time, the machine may experience a number of different pains, each one triggering different goals. Changing pains change the machine's motivation for action, concentrating its efforts on reducing the winning pain. The same mechanism that created the response to a lower level pain will govern learning how to respond to abstract higher level pains. This motivating mechanism stimulates the machine to interact with its environment and to develop its skills.

The machine continuously responds to dominating pain signals. In this search process it links new observations to desired actions. This defines and refines the concepts that it developed by observing environment. Abstract concepts are related to abstract motivations and more advanced skills developed based on these abstract motivations. Concepts that help machine to realize its goals are learned better that those, perhaps frequently observed that are not directly useful.

Many pain signals participate in development of advanced motivations and result in self-governance of the system that follows the strongest motivation. The system constantly develops its skills and abstract goals through successful interaction with the environment. By doing so, the machine not only learns complex relationships between concepts, resources, and actions; it also learns limitations of its own embodiment, and effective ways of using and developing its motor abilities. ML can be combined with curiosity based

learning, however curiosity based learning is triggered only when all the pains are reduced below a specified threshold. This typically happen in early stages of development, when the environment is simple, and the agent did not develop yet many abstract motivations.

III. COMPARISON OF ML AND RL APPROACHES IN GRADUALLY CHANGING ENVIRONMENTS

The goal of presented simulations is to examine the agent's behavior in changing environments. In our experimental setup the agent has to operate in environment where not all resources are visible nor can be used from the very beginning of simulation. Instead, the set of available resources and consequently set of actions available to the agent are gradually increasing during the interaction with the environment. At the beginning of simulation agent is able to learn only the basic dependencies between available resources. It can choose from a small set of actions. Additionally, it has a given amount of time before the environment increases in complexity introducing other resources and making new actions available. That means that environment is less complicated at the beginning. The best strategy for an agent in such situation is to gather the knowledge about environment (by creating proper, useful notions correlated with internal motivations) when it is still simple. These conditions are similar to real world situation in developmental robotics when with exploration of environment an intelligent agent is confronted with more complex world.

We compare effectiveness of learning in such "graded" environment with the one that demonstrates all its complexity from the very beginning. We want to show that when gradual emerging of new concepts and skills is correlated with increasing environment complexity learning is more effective. Our additional aim in this study is to use this environment with gradually increasing complexity to examine effectiveness of motivated learning (ML) and compare it with reinforcement learning (RL).

A. Experiments Description: Environments and Agents

Experiment setup is based on proposition presented in [14] but with some modifications. According to this concept, the environment contains few kinds of resources which may be used by agent. There are some dependencies between these environmental resources. The basic concept is that when agent uses some resource, the amount of the resource decreases and in order to replenish it the agent has to choose and perform a proper action which in turn uses another "higher level" resource. Agent can replenish every environmental resource by proper action (see column Motor in Table I) performed on proper resource. For instance if the agent is hungry it may eat food to lower this kind of pain. But by eating food it decreases available food resources and must learn how to restore them, so he can continue eating when he is hungry. In the basic experiment setup there are six kinds of resources that can be sensed and six motor actions that can be performed on any of these resources (see Table I).

TABLE I
SYMBOLIC SENSORY MOTOR PAIRS

	SENSORY	MOTOR	INCREASES	DECREASES	PAIR ID
0	Food	Eat	Sugar level	Food supplies	0
1	Grocery	Buy	Food supplies	Money at hand	7
2	Bank	Withdraw	Money at hand	Spending limits	14
3	Office	Work	Spending limits	Job opportunity	21
4	School	Study	Job opportunity	Mental state	28
5	Sandbox	Play	Mental state	-	36

We define a goal as combination of resource S and motor function M. Therefore, total number of goals is obtained as follows:

$$|G| = |S| \times |M| \qquad (1)$$

where:
$|G|$ - total number of goals
$|S|$ - total number of sensed kinds of resources
$|M|$ - total number of motor functions

Moreover the environment is hostile to the agent. This means that there are only small amounts of all kinds of resources and while consuming them the pain signal P increases. The aim of agent's learning is to learn all relationships between different resources in order to be able to control their amounts and thanks to this be able to minimize internal pain signal value. Specifically, the agent should learn which actions to perform in order to replenish the resource which is needed at this very moment. The following function describes the probability of finding resources in this experiments setup:

$$f_{ci}(k_{ci}) = e^{-\frac{k_c}{\tau_c}} \qquad (2)$$

where:
τc – scaling factor that describes a resource declining rate
kc – number of times a resource was used
This kind of environment is dynamic which means that its state changes as a result of actions performed by the agent. The environment state is observable. In our experiments we use environments with different number of resources and available motor actions: from 6 to 18. The more resources are available in environment the higher the environment complexity. This kind of environment where all the resources are available from the very beginning of simulation is called the "normal" environment. Another kind of environment, called "graded" environment gradually increases the set of available resources and actions that the agent can perform.

Our computational model consists of two components: the environment module (described earlier) and the agent module. We used two kinds of agents: one based on motivated learning approach and another one based on reinforcement learning approach. Foundations of ML approach are described in [14] in details. The RL approach was implemented through TD-Falcon algorithm [15]. It is a generalization of Adaptive Resonance Theory – a class of self-organizing neural networks – that incorporates temporal difference methods for real time reinforcement learning (TD-Falcon stands for Temporal Difference – Fusion Architecture for Learning, Cognition, and Navigation). This algorithm learns the value functions of the state-action space using temporal difference methods, and then uses them to determine the optimal action selection policy.

B. Normal vs. Graded Environment

The first experiment illustrates the advantage of a gradual increase in the environment complexity for ML machine development. Thus as machine develops new concepts and learns how to deal with them, it is prepared to handle more advanced and complex situations.

In this experiment we have used two kinds of environments - "normal" and "graded". We have conducted simulations in environments with: 6, 10, 14 and 18 different resources. That means that there were from 6 to 18 hierarchy levels (each one representing different resource). The results obtained are illustrated on Fig. 2 which shows the number of iterations needed for a ML machine to learn the environment's complexity in "normal" and "graded" environments.
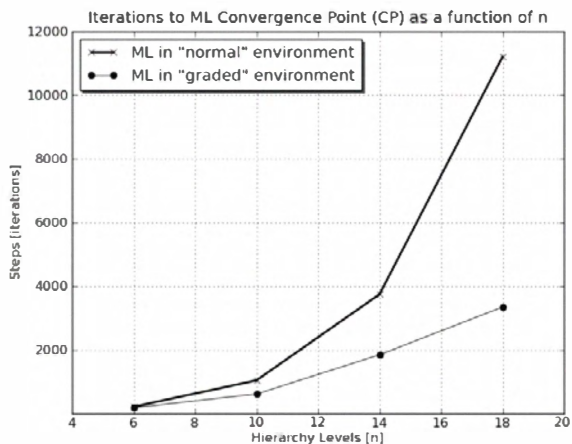
Fig. 2. Results for ML agent in environments with constant complexity ("normal") and with gradually increasing complexity ("graded").

The outcome of this experiment shows that ML agent can learn more effectively in environment with gradually increasing complexity (called "graded" and depicted with the thin line) than environment with constant complexity (called "normal" and depicted with the thick line). For example ML agent operating in "graded" environment needed about 3850 iterations to learn all environment dependencies but in "normal" version of it agent needed over 11000 iterations. For the bigger environment this difference becomes more

significant. This means that in the case of gradually changing complexity ML agent has enough time to learn some basic motivations (corresponding to internal dependencies between objects in the environment) before the environment becomes more complicated.

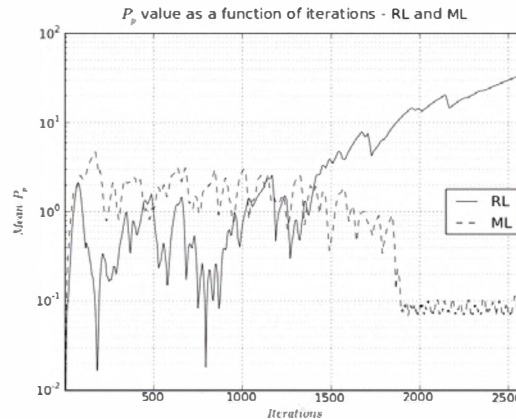### C. Reinforcement vs. Motivated Learning

The second experiment compares effectiveness of development between ML and RL based agents. In this simulation we have used environment with gradually increasing complexity. Obtained results are depicted on Fig. 3.
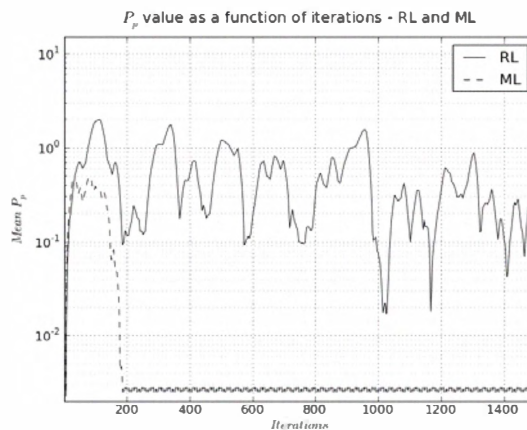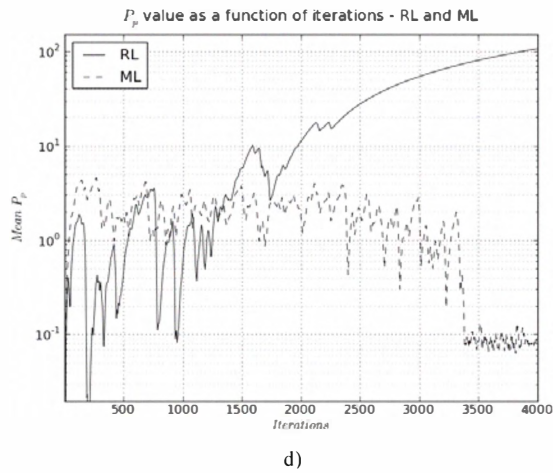
a)
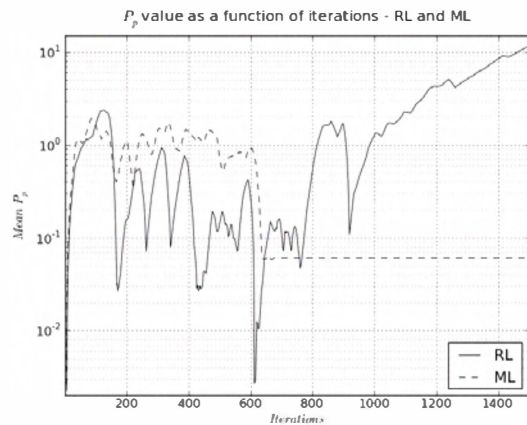
b)

c)

Fig. 3. Moving average of P_P value as a function of number of iterations during computational experiment in "graded" environment with: a) 6, b) 10, c) 14, d) 18 levels of hierarchy

In the case shown in Fig. 3a) where the environment complexity is relatively small (6 levels of hierarchy), ML based agent experiences similar internal primitive pain signal P_p as RL based agent. However, unlike the RL agent, ML agent converges quickly to a stable performance. Then after reaching the convergence point (having learned all dependencies in the environment) its mean P_p signal is from 10 to 250 times smaller than RL agent's. The ratio of P_p signal values for both methods is depicted on Fig. 4a).

In more complicated environments (with 10, 14, 18 levels of hierarchy) the situation is a little bit different. Before reaching the convergence point, P_p signal of RL agent is sometimes lower than P_p signal yielded by ML agent. However, after an initial success, RL is not able to learn complexities of increasingly hostile environment and diverges in all these cases. However, the ML agent is still able to converge (Fig. 3b), c) and d)).
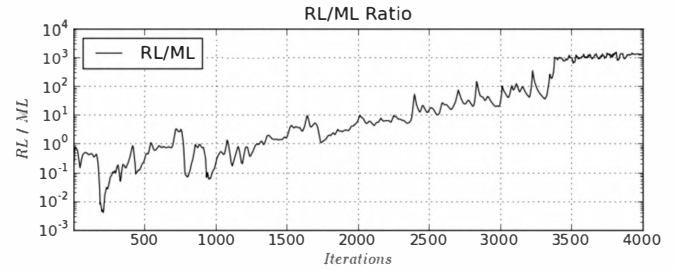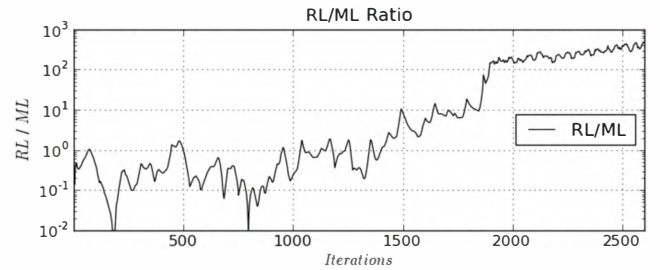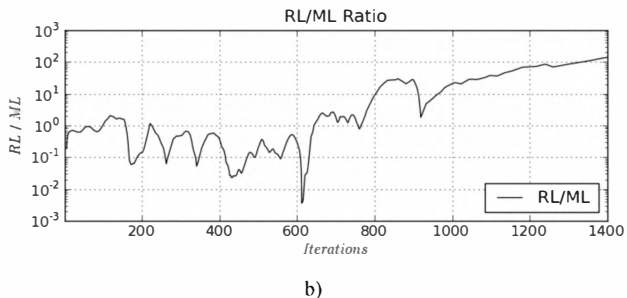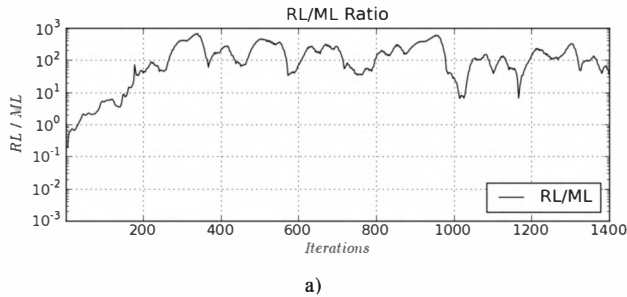

a)

b)

c)

d)

Fig. 4. RL/ML P_P ratio for : a) 6, b) 10, c) 14, d) 18 levels of hierarchy
.

The ratio between values of P_p signals reaches values form 100 to 1000 in environments with 10 and 18 levels respectively. That means that agent based on reinforcement learning approach is not able to adjust to the environment with gradually changing complexity.

These examples illustrate the advantage of effectiveness of ML over RL approach. Consequently, we can conclude that while RL approach is not optimized for environments that are dynamically changing, ML based approach works well in this kind of environments.

D. Summary of Experiments

In presented experiments we have demonstrated some aspects of proposed motivated learning approach. Our goal in this study was to examine effectiveness of learning in special kind of dynamical environments. The main problem in acting in such environments is that agent has to learn all dependencies between different factors and know which action to take in specific situation.

We used two types of the environments. In the first one, called "normal", all resources are available to the agent from the very beginning of the simulation, while in the second one, called "graded", the set of resources which agent can observe and use are gradually increasing. The main feature of this environment is that it can nurture development of skills (requiring gradually more complex concepts) of an agent interacting with it. As experiments revealed ML agent can learn more effectively in environment with gradually increasing complexity than in the environment with constant complexity. This means that it is useful to learn basic notions and skills in relatively simple environment and then use them in more complicated situations.

Our other aim in this study was to examine the ML agent's learning effectiveness in comparison to RL agent's. As experiment revealed, ML agent's learning effectiveness were much better than those of RL agent. ML based agent was able to converge to stable solution while RL based agent cannot do that in this kind of environments.

IV. Conclusion

A bottom-up connectionists approach delivers sensory information used in building object representations. In unsupervised learning this corresponds to clustering without labels. Top-down approach, characteristic to symbolic manipulation, uses the sensory information received to direct machine's actions. In the motivated learning these two functions are combined into one scheme.

The emergence of new concepts that are obtained from bottom-up representation building is supported by the functional use of these concepts to benefit the machine. This functional use represents a top-down control over the machine, yet this control emerges together with supporting object representations through a resonance between what is being observed and what machine wants to do with the observed information.

Thus top down decisions are very much a function of proper understanding of the bottom-up representation in the learning cycles that reward proper interpretation of the input data, and the internal states of the machine represented by the competing pain signals. Proper use of the observed objects, tools and resources, defines their practical meaning to the machine that utilizes them. At the same time, stability of the observed representations provides a basis for mental development of abstract motivations and goals, yielding mental processes like planning, thinking and top-down control.

The proposed ML method shows how mental development stimulates learning of new concepts and at the same time benefits from this learning. Thus the method is suitable for machine operation in developmental robotics.

Comparison to reinforcement learning indicates weakness of the reinforcement learning method in the environments with increasing complexity, even if such increase is gradual and the environment provides ample opportunities to learn in the early stage of machine development.

References

[1] B. Bakker and J. Schmidhuber, "Hierarchical Reinforcement Learning with Subpolicies Specializing for Learned Subgoals," in M. H. Hamza (Ed.), Proceedings of the 2nd IASTED International Conference on Neural Networks and Computational Intelligence, NCI 2004, Grindelwald, Switzerland, pp. 125-130, 2004.

[2] Legg, S. & Hutter, M. (2006). A Formal Measure of Machine Intelligence," *Proc. 15th Annual Machine Learning Conference of Belgium and The Netherlands,* pp.73-80.

[3] A. G. Barto and S. Mahadevan, "Recent advances in hierarchical reinforcement learning," Discrete Event Systems, Special issue on reinforcement learning, 13, pp. 41–77, 2003.

[4] R. Parr and S. Russell, "Reinforcement learning with hierarchies of machines," In *Advances in Neural Information Processing Systems: Proceedings of the 1997 Conference*, Cambridge, MA, 1998.

[5] T. G. Dietterich, "Hierarchical reinforcement learning with the maxq value function decomposition," *Journal of Artificial Intelligence Research*, 13, pp. 227–303, 2000.

[6] W. Schultz, "Getting Formal with Dopamine and Reward," *Neuron*, Vol. 36, pp. 241-263, 2002.

[7] J. Schmidhuber, "Curious model-building control systems," *Proceedings Int. Joint Conf. Neural Networks*, Singapore, vol. 2, pp. 1458–1463, 1991.

[8] P-Y Oudeyer, F. Kaplan, and V. Hafner, "Intrinsic motivation for autonomous mental development," in *IEEE Transactions on Evolutionary Computation*, 11(2), pp. 265–286, 2007.

[9] F. Kaplan and P-Y Oudeyer, "In search of the neural circuits of intrinsic motivation," *Fontiers in Neuroscience*, vol. 1, Jan 2007.

[10] S. Roa, G. J. M. Kruiff, and H. Jacobson, "Curiosity-driven acquisition of sensorimotor concepts using memory-based active learning," *Proceedings of the IEEE Intl. Conf. on Robotics and Biometrics*, pp. 665-670, 2008.

[11] X. Huang, J. Weng. Inherent Value Systems for Autonomous Mental Development, International Journal of Humanoid Robotics, 4(2), 407-433, 2007.

[12] D. Vernon, G. Metta, and G. Sandini, "A survey of artificial cognitive systems: Implications for the autonomous development of mental capabilities in computational agents," *IEEE Transactions on Evolutionary Computation*,11(2) , pp. 151-180, 2007.

[13] A. Barto, S. Singh, and N. Chentanez, "Intrinsically motivated learning of hierarchical collections of skills, *Proc. 3rd Int. Conf. Development Learn.*, San Diego, CA, pp. 112–119, 2004.

[14] J. A. Starzyk, "Motivation in Embodied Intelligence" in Frontiers in Robotics, Automation and Control, I-Tech Education and Publishing, Oct., pp. 83-110, 2008.

[15] A.-H. Tan, L. Ning, X. Dan, "Integrating Temporal Difference Methods and Self-Organizing Neural Networks for Reinforcement Learning with Delayed Evaluative Feedback," IEEE Transactions on Neural Networks, Vol. 9, No. 2 (February 2008), pp. 230-244, 2008.