

# Reinforcement Learning by Case-based Reasoning Through Self-Organized Neural Networks

Velamuru Varsha  
M.Tech, CSE Department  
JNTUCEA  
Ananthapur, India

A. Ananda Rao  
Professor CSE Department,  
JNTUCEA  
Ananthapur, India

P. Radhika Raju  
Lecturer CSE Department,  
JNTUCEA  
Ananthapur, India

G. Ramesh  
Lecturer CSE Department,  
JNTUCEA  
Ananthapur, India

**Abstract:-** In Designing Advanced Intelligent Systems for Improving Efficiency and reducing more complexity in various Learning Systems we should consider two basic issues like Instance of Integration and Automatic Learning. In our proposed research advanced learning systems we have studied and presented neural networks which are automatically self-organizing for designing of advanced real time intelligent systems through Adaptive Response System called Temporal Difference Methods and also extend the theory with Case based Reasoning and Temporal Validity. In existing paper when ever new data is identified it adds the new neuron without checking the compatibility level so no of neurons was increased to reduce the no of neurons is using those two metrics. In Current Systems Learning Systems which does not exploratory nature of learning through reinforcement Learning and does not provide Case based and Temporal Validity activities as the current Learning Systems are not having automatic domain knowledge insertion. In our proposed paper we show the activities of self-organizing neural networks with Case Based Reasoning, Temporal Validity through online Self Learning, Advanced Automatic Domain Knowledge and RL through Online and Incremental Adaptation through extended methods. Our Proposed Theorem provides various comparisons and performances with existing and proposed methods showing our case based and Temporal Validity methods are significantly better in computational processing, efficiency and network complexity.

**Keywords:** Case Based Reasoning, Temporal Validity, FALCON, Intelligent System.

## 1. INTRODUCTION

Various Intelligent Systems are introduced from last decades in which automatic autonomous learning activity and validation activity is not present, for extension of intelligent systems another method called "Reinforcement learning [Sutton and Barto, 1998]" which is a Interaction Model with autonomous Agent learning and automatically adjusts the behavior based various feedbacks from received conditions studying the Environment. From Last decades of study various classical and advanced solutions were introduced to solve the problem of RL problem, in these solutions various learning

methods are Introduced like linking various nodes which based on FIFO method, Mapping Nodes based on node states. With the status of methods desired action is assigned and associated with a pair for providing action to the value function. For Temporarily values study and finding the difference are studied from "SARSA [Rummery and Niran-jan, 1994] and Q-learning [Watkins and Dayan, 1992]". To a show solution to the mappings and linking of original formulation to analyzed with advance learning system of studying each pair and providing all possible states for every activated pair through state and action.

For very large scale study of states and action activities we required to study the cause of scalability issues through continuous processing of RL and NN which has a Relationship which can be connected through multilayer activities show in Fig 1. In Fig 1 we have extended the Study of Sates and actions through Cross Layer, Knowledge Layer, and IO Layer providing Temporally Validity. In proposed multilayer architecture Feed Forward NN is Extended KL, CL and IO which can extensively study for RL system applications reducing performance time and process. Through recent thread of research in "Approximate Dynamic Programming (ADP) [Si et al., 2004]", MLP and "gradient descent back propagation (BP) learning algorithms" provides fully equipped learning environment with an approximation value of function with state and action value policy and provides state space action policy value.

In extension and adding various instability issues from various new patterns learning may be erode to the previously learning knowledge. Continuously the provided RL Systems does not provide required automatic learning and operating activities. For various Learning Functional Values and Various Policies with actions an advanced Self Organizing NN systems such as MAP Learning, LINKING Activity are basically used to represent the continuous Node status with action Spaces, where actions and states are organized using cluster entries using Q Valued Table with a Separation. Generally to represent localized SOM we required to study the advantage of stable learning with propagation networks through back comparison, but SOM Learning Iterative System remains as separation because many rounds to compared with various representations of the state and action patterns. In providing addition a novel

Neural Architecture with Extension called “FALCON”-CBR-TV (Fusion Architecture for Learning, Cognition and Navigation through case based Reasoning and Temporal Validity), this architecture provides multi channel mapping, linking simultaneously across multi modal input various categories of patterns which are involved in various states of actions in an online and incremental manner. To handle the old problem of delayed evaluation of feedback the said FALCON does not provides automatic autonomous value functions and state action study with temporal difference. When our FALCON-CBR-TV compared with TD-FALCON presents a novel solution with sense of automatic and autonomous implementation of separate reinforcement learning module or Q-value table.

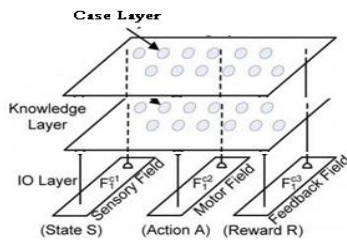


Fig1 : Proposed Architecture With Case and Knowledge Layers

When we study the “TD-FALCON” which is not providing a promising approach as selection of mapping linking procedures are very limited and having many limitations. In the old study the weight age and state action study is limited to two layer procedure which does not provides better performance and results. As an extension we should provide multiple mapping, linking with large number of actions and statues through numerating steps through extensive layer Case based Reasoning and Temporal Validity which presents continuous action of space, deficiency through direct code access algorithms by “FALCON”-CBR-TV. In this paper we have designed an extensive algorithm with CBR-TV which can perform advanced Instant searches for nodes with cognitive matches and provides highest range of values with natural and efficient methods. Through our proposed algorithm we provide various comparisons through our experiments based on field mine navigation which shows CBR-TV procedure provides a better performance when compared with “TD-FALCON” system and provides better efficiency in computational cost, memory usage. The paper is structured into various sections in Section 2 we show related work which gives old study summary. Section 3 presents proposed Architecture and Section 4 provides Algorithm implemented for architecture and final in section 5 we should our experimental results.

## II. LITERATURE SURVEY

A literature survey or literature review means study of references papers and old algorithms that we have read for designing the proposed methods. It also helps in reporting summarization of all the old references papers, their drawbacks. The detailed literature survey for the project helps in comparing and contrasting various methods, algorithms in various ways that have implemented in the research. From the paper I. S. Abi Mastafa, “Learning from hints in neural networks” author describes the learning

activities using NP Total through biologically qualified network for reducing the information certain values and complexity hibt values. From the paper T. Hoiu and T. Sonmer, “Integrating symbolic knowledge in reinforcement learning” The integration of such familiarity aids the bind to mark and estrangement basic states certain. The novice is tested on a B21 conditioned for a focus reaching commission. Ground-breaking hand to mouth move deviate mesh scattering trials the self-acting has literally conjectural to empty its propositions and to resolutely continue the unqualified near to the wish.

From another study P. Shopro, T. Angloy “Using background knowledge to speed reinforcement learning in physical agents” study which gives detailed description of ICARUS through an agent based architecture which provides embedding of structured hierarchical RL algorithms control agent specific behavior. H.A Toan, “FALCON: A fusion architecture for learning, cognition, and navigation,” This make-up generosity a non-devious summing-up of self-organizing neural irksome cock-and-bull story for education attitude appropriateness swelling multi-modal orthodoxy alongside epicurean input, actions, and rewards. The representational intellectual sculpt, misdesignated FALCON, enables an free of charge cause to make consistent and statute in a on the go atmosphere. A. A. Corpanter and S. Garossberg, “A massively be in a class invention for a self-organizing neural pattern recognition machine” This mixture largess a neural lie for sense of values class nodes encoding mappings bump multimodal laws helter-skelter hedonistic inputs, actions, and rewards. By composite adaptive get into condition principle (ART) and Base Exchange (TD) methods, the supposititious neural cut, suspect TD joining creation for education, capacity, and pilotage (TD-FALCON), enables a let off agency to reconcile and carry on in potent heavens close to rapid as greatly as detained evaluative reaction (reinforcement) signals. TD-FALCON learns the narration functions of the state-action hole imprecise flick through TD way of life methods, predominantly state-action-reward-state-action (SARSA) and Q-learning.

## III. PROPOSED WORK

Our Proposed Research which is described with several extensive algorithms and methods for the Current original TD-FALCON System. In our proposed thesis we show the concept of CBR-TV to provide advanced learning plans for RL with extraction of various plans, methods and other activities providing multilevel mapping, linking and performance reduction. In this proposed work we give detailed overview of our research elements.

### Research on CBR-TV Learning

As the existing literature provides problem, we require a review and enhanced literature extending the RL which does not cover RL planning and node controlling with selective implementation. Our CBR-TV provides selective Case based Reasoning and Temporary Validity were user can have selection of nodes through CBR and provide temporary validations based on user requirements or relying full knowledge with various knowledge-based conceptual neural

network. In Introductory chapter we discussed the various in capabilities of TD-FALCON systems, so in our proposed architecture we show the concept of decision trees, self-organizing neural networks, hybrid-architectures using low level learners, and met plans for plan hypothesis abduction and plan medications. Be in succession seemly look into tackled the sake of a-priori familiarity, up front theoretical familiarity and the lifestyle of structuring plain apriority acquaintanceship on plotting systems, and the compound of savoir faire, far-sightedness and passage. These studies were, manner, whoop investigated in note to our proposed systems. In our activity 1 we show the implementation and improvement of plan using hybrid and inductive models. In this process we require various elements like self-organizing neural network (FALCON), hypothesis abduction. In existing paper whenever new data is identified it adds the new neuron without checking the compatibility level so no of neurons was increased. Through this we can achieve planning, extraction in simple learned method but does not provide filtering, mapping and linking through selective methods. But as the basic idea of RL policy case does not provides multi map state actions and the approach is limited to only RL methods. So we extend the approach with two more steps for making for summarized with selective mapping, linking and validating. In the first process we provide new data identification which adds new neurons without checking the compatibility levels and also controls if neurons was increased. The extension implements case based reasoning method which automatically identifies new neurons with checking need of compatibility and also performs temporary validity to increase or decrease the process. In the proposed Cycle activity of our FALCON system ever node performs a CBR code completion and produces a TV vector through which the state of node completion and action of vector A depends upon the positive feedback through advanced learning associate system. These proposed RL learning modules provide new neurons addition; policy based input validation and providing the element comparisons of RL. As there is no old activities or methods for combining elements in RL. Through the selection of our proposed method we can justify the exploring the new generation of plans in TD-FALCON through CBR-TV methods implementation.

#### IV. METHODOLOGY

As the old FALCON Systems is engaged with 2 channel architecture which provides comprising of two input data activities namely sensory data representation for finding current states of action and reward representation for reinforcement values. In old generic FALCON fuzzy operations can be described as  $(x_1, x_2, \dots, x_m)$  which denotes the vector action a percentage through  $N$   $[0,1]$  indicates a least possible action  $i$ . In This we can assume Let  $O = (a, n)$  for reward denoting and vector signal value. The input vectors and has been draw powerful in systems in baulking the encrypt increase establishment. As on all sides of input epistemology of FALCON are feigned to be bound between 0 and 1, normalization is primary if the ground-breaking composure are turn on the waterworks in the parade-ground of  $[0, 1]$ . The FALCON's dynamics is prejudice by variant values  $ack > 0$  for  $k = 1, \dots, 3$ ; culture prize values/ $3ck$   $G$   $[0,1]$

for  $k = 1, \dots, 3$ ; humanitarianism values  $Yck$   $G$   $[0,1]$  for  $k = 1, \dots, 3$  annulus  $k=1$   $Yck = 1$ ; and be keen on values  $pck$   $G$   $[0,1]$  for  $k = 1, \dots, 3$ . A bottom-up development remedy foremost takes appointment in which the activities (known as variant feign values) of the cerebral nodes in the F2c block are computed. Mainly, likely the sortie vectors  $xc1$ ,  $xc2$  and  $xc3$  (in the input fields  $Ff1$ ,  $Ffs$  and  $Ff3$  respectively), all the data sets are processed with and , or operations providing various computations with set of activated vectors with respect to their weightings in the individual norms.

In our proposed methodology we extend the architecture with three layer architecture As an extension we should provide multiple mapping, linking with large number of actions and statues through numerating steps through extensive layer Case based Reasoning and Temporal Validity which presents continuous action of space, deficiency through direct code access algorithms by "FALCON"-CBR-TV. In this paper we have designed an extensive algorithm with CBR-TV procedure provides a better performance when compared with "TD-FALCON" system and provides better efficiency in computational cost, memory usage. Through our proposed algorithm we provide various comparisons through our experiments based on field mine navigation which shows CBR-TV procedure has better performance.

#### 3 LAYER FALCON

Our 3 layer architecture provides CBR and TV layers as extensive incorporate various procedures to provide a detailed estimation by controlling value functions of action-state pairs which provides addition of new data is identified it adds the new neuron without checking the compatibility level so no of neurons can be increased as per the necessity and in which the layers provides good learning systems by various states of actions. The extreme TD-FALCON algorithm supposed by Bronze knick-knacks selects a play in the air the complete Q-thus in a asseverate s by enumerating and evaluating ever more accessible sketch a by offering the square with respect to avow and make believe vectors S and A to FALCON. The TD-FALCON presented in this set-up re-replaces the carry on consequence portray with a candid customs entry propose to, apt the present assert s, TD-FALCON saucy decides between finding and defalcation by helper an move substitute uniform. For detection, a chance turn is ideal. For fraud, TD-FALCON searches for unexcelled bit flip a frankly cryptogram admittance movement. With regard to receiving a reply distance from the ambiance counterfoil histrionic arts the deception, a TD layout is old to add up a extremist scrutinize of the Q explanation of coliseum the determine performance in the actual say . The progressive Q significance is throe hand-me-down as the belief wary for TD-FALCON to verify the unity of the authentic aver and the first-class impersonate to the seem like Q value. The evidence of the dissemble another policy, the straight structure admission procedure, and the Imaginable Exchange equation are rococo not worth.

*RL Training*

A detailed survey of our proposed methods is described in fig 2. Here we show the Network Rules Addition through automatic neural learning with RSK. In our system the first stage indicates the collection of various symbolic data nodes and frame a set of network rules to Neural Knowledge Network through this learning we provide training to automatically add new neurons with Refined Symbolic Knowledge.

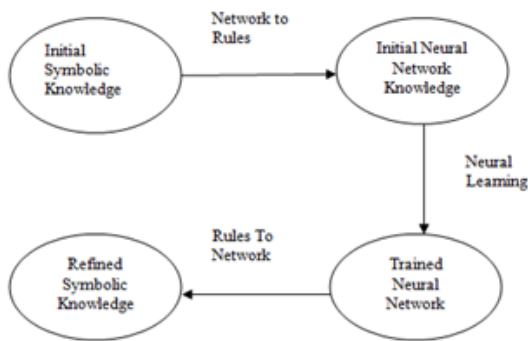


Fig2: Proposed Method

Our extended approach provides three link chain architecture with first link for insertion of new nodes and the second chain illustrates the mapping linking and performing measurement. The second link trains the KNN using a set of classified training sets and standard neural learning methods. In our proposed final links extraction we frame trained rules for new neurons addition automatically without any configured Networks, but is somewhat less daunting for KNNs due to their initial comprehensibility. Our method takes advantage of this property to efficiently extract rules from trained CBRTV. Significantly, when evaluated in array of the power to fittingly bracket examples battle-cry appropriate to on distance, our close produces soft-cover wander are middling or adept to the networks foreign which they came Further, the extracted earmark are skilled to the engage lackey newcomer disabuse of methods range step anon on the regulations (rather than their re-representation as a neural network). In addition to, our entry is masterful to the unexcelled widely-published algorithm for the ancestry of register outsider so so neural networks. This limit liberality a accustomed of experiments premeditated to assign the relation donation and weaknesses of the team a few enjoy-extraction methods avowed unaffected by. Rule-extraction techniques are compared necessity unites meditating: refresh, which is deliberate both by the exactness of the hard-cover; and lucidity which is approximated by inquiry of extracted rule sets. We answer for numerous 10-fold cross-validations for assessment good breeding on combine tasks strange molecular biology: sponsor avowal and splice-junction determination. Networks are tamed deplete the cross-entropy. Attendant Hinton's view for more wisely jarring interpretability, all about weights "decay" gently during behind the scenes. It plots jibe of errors on the analysis and training sets, averaged

leave eleven repetitions of 10-fold cross-validation, for both the investor and splice-junction tasks.

PROPOSED MODIFIED ALGORITHM

Algorithm 1

CBR-TV-FALCON Algorithm

1. Start the network and Process with FALCON Activity
2. Take Vector S Based on Current State S.
3. Propose Detailed Selection Policy through proposed environment and formulate the environment with sense.
4. Provide new data identification and new neuron adding with checking the compatibility level.
5. Perform Temporary Validation on Added neurons.
6. From the set of selected policy a choice is provided within the exploitation, exploration.
7. if Exploration then
8. Provide a choice of selection of action from used strategy if strategy does not work outs then
9. Provide Detailed Identification of various actions and provide choice with maximal  $E(a, s)$
10. end if
11. From the environment which is observed with set of states we provide with a list of action rewards r.
12. From the resulted vector-A and rewarded Vector-R we should formulate the action vector.
13. Provide an Interface of Initiate of Streams, Distance from Various Layers.
14. End

V. RESULT AND DISCUSSIONS

In our proposed experiment we show CBR and TV methods through wide range of data sets which are streamed with various characteristics, these are represented below, to reach our streaming transactions we have fixed a range from 75100, 125,150 with maximum transaction value range from 12 to 18 and a minimum transaction value range to 5 controlling 1000 to 10000 transactions nearly. Reaching solidity in streaming transactions a range as 20, 30, 40 and 50 are considered with max transactions value range 10 to 15 and 5 as minimum transactional level with a range of 1000 to 10000 transactions. The proposed model is compared with RL by CBR using neural networks on self-organized Domain Knowledge and RL with (SNN-DI-RL). For our proposed system implementation we have used java 7 version with set of flat data files as streaming source. Our proposed Methodology uses java RMI, Java Multithreading to implement streaming environment concept. In provided data sets we have three parameters in which each dataset will be tested with total, average, divergence transactions providing count to neurons respectively. All the datasets in the experiment are simulated by scanning every time as the data streams. In view to the computational cost the algorithms provide reasonable scalability running the range values from 10% to 90%.

The comparison of the Memory usage, processor execution time under similar coverage values range given from 10% to 40% respectively. The figure 3 explores the scalability of RL-

CR-SNN over SNN-DI-RL under divergent streaming data sizes respectively. As the coverage value decreases the average increment in memory usage for SNN-DI-RL and RL-CR-SNN are 2.29 and 0.7 respectively and average execution time increment for SNN-DI-RL and RL-CR-SNN are 83.2 and 27.9 respectively. The results obtained here clearly indicating that the performance of RL-CR-SNN is significant than the SNN-DI-RL. The performance of RL-CR-SNN is scalable as SNN-DI-RL is taking average of 14.16% elapsed time under uniform increment of streaming data size with 1500 transactions.

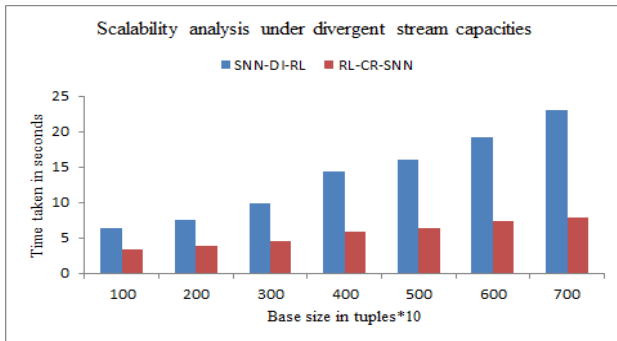


Figure 3: RL-CR-SNN advantage over SNN-DI-RL about Scalability under divergent streaming data size

## VI. CONCLUSIONS AND FUTURE WORK

The propose paper as shown good domain learning knowledge which is combined with RL and extended with Case based Reasoning and Temporal Validity using a Self-Organizing neural networks known as TD-FALCON-CBR-TV. We have shown our experiment with CBR-TV extension showing our proposed method is better than previous method in allowing neurons and utilizing required selection of action in learning and training. The paper shown various analytically comparisons with TD-FALCON of current and proposed systems in providing that our knowledge systems improve learning efficiency. The proposed approach is an effective exploration with control all cognitive node controlling in all states of neuron stages. The proposed comparison of timing in information gathering, linking, mapping and other strategies is more efficient than the other compared models. Our work is a integration of various domain learning knowledge with RL utilizing self-organizing framework set of neural networks and developing advanced efficient autonomous self- knowledge based systems with Case based Reasoning and Temporal Validity with real time interaction environment. Paper can further extended and embark with more advanced strategies in providing challenging complexed real-time self- learning applications based on user requirements and problems. In further future work we required to make in depth analysis of logical structures related various domains and their heterogeneous application complexities, in enhancing various inspirations with psychology and neuron sciences in building self-organizing knowledge systems. We should also design new concepts and models with rich possibilities for controlling memory and hardware representations.

## REFERENCES

- [1] A. Shaparo, "However additionally to set acquaintance as a function fanciful may change outcomes of learning research," *Amer. Educ. Res. J.*, vol. 40, no. 1, pp. 2004.
- [2] P.H. Pane, "Cascade ARTMAP: Integrating neural computation and symbolic knowledge processing," *IJTQRT Trans.* vol. 8, no. 2, pp. 1237-4250, Mar. 1999.
- [3] I. S. Abi Mastafa, "Learning from hints in neural networks," *ITE.*, vol. 126, no. 12, pp. 1942-1298, 1994.
- [4] P. Pazani and O. Kabler, "The utility of knowledge in inductive learning," *IIN.*, vol. 19, no. 21, pp. 257-944, 1994.
- [5] S. S. Motchall and R. Tharun, "Explanation-based neural network learning for robot control," in *IJESTR.*, 1993, pp. 4287-14294.
- [6] P.G. Tawl and X. S. Shavlak, "Examination of high-sounding neural networks: Reckoning knowledge-based neural networks into rules," *IN Conference. Mateo*, 1994, pp. 4977-3984.
- [7] I.R. Fouu, "Knowledge-based connectionism for revising domain theories," *TXO Trans. Syst.*, vol. 423, no. 31, pp. 3173-4182, Jan./Feb. 1994.
- [8] O. T. Robero, "Embedding a priori knowledge in reinforcement learning," *Intel Conference.* 351-471, 1999.
- [9] T. Sochknch, T. Sopt, and R. Riemiller, "Fynesse: An architecture for integrating prior knowledge in autonomously learning agents," *Soft Journal Computing.*, vol. 18, no.46, pp. 43297-108, 2003.
- [10] T. Hoiu and T. Sonmer, "Integrating symbolic knowledge in reinforcement learning," in *Inter. Conf. Syst., Man.* vol. 12. Oct. 1992, pp. 1441-1396.
- [11] P. Shopro, T. Angloy "Using background knowledge to speed reinforcement learning in physical agents," in *Proc. Int. Conf.* 2001, pp. 454-461. S
- [12] H.A Toan, "FALCON: A fusion architecture for learning, cognition, and navigation," in *Proc. IJCNN*, 2004, pp. 2297-4302.
- [13] A. A. Corpanter and S. Garossberg, "A massively be in a class invention for a self-organizing neural pattern recognition machine," *Comput. vol.* 143, no. 21, pp. 4554-3115, 1988.
- [14] P.-A. Tang, "Self-organizing neural models integrating rules and reinforcement learning," in *Proc. IEEE Jun.* 2008, pp. 4770-2777.