IJCTA, 10(8), 2017, pp. 299-312 ISSN: 0974-5572

Predictive learning model in Cognitive Radio using Reinforcement Learning

Sharada Tubachi*, Mithra V. and A.V. Kulkarni*****

ABSTRACT

In the current years, the scenario of wireless communication is changed radically, resulting in advent of multiple innovative technologies such as Wi-Fi, WIMAX, and cognitive radio. Cognitive Radio is a special type of software defined radio and network technology that can detect available unused channels in a wireless spectrum automatically and change its transmission parameters enabling more communications to run simultaneously and also improve radio operating behavior without interfering with primary users. Cognitive radio will have self-adaptability to adjust its operating parameters to perform the communication which is possible only because of it's the cognition part called as the intelligence concentrated part of cognitive radio where different intelligent techniques are applied. One of the techniques is reinforcement learning technique. Our proposed learning technique is capable of interacting with uncertain environment and it does not require any previous knowledge of the environment still it can achieve high performance.

***Index Terms*:** Artificial Intelligence, Cognitive Radio, Future Wireless Communication System, Reinforcement Learning.

1. INTRODUCTION

A basic problem facing in the future wireless systems is where to find appropriate spectrum bands to satisfy the demand of future services. While all of the radio spectrum is allotted to different services, applications and users, observation show that usage of the spectrum is actually quite low. To overcome this problem and improve the spectrum utilization, cognitive radio concept has been developed. Wireless communication, in which a transceiver can detect intelligently, communication channels that are in use and those which are not in use are known as Cognitive Radio and it can move to unutilized channels. This makes possible use of available radio frequency spectrum whilst minimizing interference with other users. CRs must have the capability to observe, learn and assign their wireless transmission according to the surrounding radio environment. The application of Artificial Intelligence approaches in the Cognitive Radio is very gifted since they have a great importance for the implementation of Cognitive Radio networks architecture. There are several artificial intelligence techniques available such as artificial neural networks, reinforcement learning, hidden Markov model, fuzzy logic etc. Reinforcement learning is a type of unsupervised learning and online artificial intelligence technique that improves system performance using simple modeling. Since there is no external teacher to supervise learning process, the agent learns about the environment by itself.

Reinforcement learning (-RL) has been applied in CR so that the SUs can observe, learn, and take optimal actions on their respective local operating environment. For example, a SU observes its spectrum to identify white spaces, learns the best possible channels for data transmissions and takes actions such as to transmit data in the best possible channel. Examples of schemes in which RL has been applied are

* PG scholar DYPIET SPPU Pune, *Email: isharada416@gmail.com*

** Assistant professor DYPIET, *Email: mithrav@rediffmail.com*

*** Dean R&D DYPIET, Pimpri, *Email: anju_k64@yahoo.co.in*

dynamic channel selection, channel sensing, routing, security enhancements, energy efficiency enhancements, channel auction, medium access control and power control.

2. COGNITIVE RADIO

Now a day's wireless and radio communications are used rapidly and its area has been developed enormously and intended to be more developed in the nearby future. Nearly 3.5 billion devices are using the wireless technologies and the number is expected to increase 50 percent more than today. We know that the gadgets like laptop, automobiles, TV, cell phone and tablet PCs which use this wireless technology are very smart in performing their function which also help us in our daily routine, and make our work very easy. This usage of these many devices make for more demand of spectrum [11]. A basic problem facing in the future wireless systems is where to find appropriate spectrum bands to satisfy the demand of future services. While all of the radio spectrum is allotted to different services, applications and users

As a first step towards to solve the spectrum scarcity problem is opportunistic spectrum access (OSA), which allows unlicensed users to exploit unutilized licensed spectrum, in such a way that limits interference to primary (licensed) users. Fortunately, technological advancement in cognitive radios, which have recently been known as the key enabling technology for realizing OSA [12]

However, some of the known and most popular spectrum bands are allocated and still not utilized fully, most of the time usage of the spectrum is actually quite low. The demand for wireless communication introduces efficient spectrum utilization challenge. To address this challenge, cognitive radio (CR) is emerged as the key technology; which enables opportunistic access to the spectrum. Cognitive radio (CR) becomes a hot research topic in wireless communication domain owing to its advantages of dynamic spectrum access and intelligent adaptation to environment [10].

The main purpose of this cognitive communications is to examine the spectrum utilization issues related with primary and secondary user (SU) systems sharing common spectrum [6]. Cognitive radio (CR) is the future generation wireless communication system that facilitates unlicensed or Secondary Users (SUs) to explore and use underutilized licensed spectrum (or white spaces) owned by the licensed or Primary Users (PUs) in order to improve the whole spectrum utilization. The CR technology improves the available bandwidth at each SU, and so it enriches the SU network performance [1]. The primary factor that helps CR to perform better in various situations is its operating frequency, in which cognitive radio is capable of changing the frequency and act according to it. Based on the information it gets from its surroundings it determines the most appropriate frequency is selected and communication is done according to the frequency available and frequency determined. The parameters in CR like capability and configurability makes the radio to act as smart and hence the name cognitive radio.

The definition of cognitive radio suggested by ITU-R is: 'a radio system employing a technology, which makes it possible to get knowledge of its operational environment, policies and internal state, to dynamically modify its parameters and protocols according to the knowledge obtained and to learn Since it clears that the basic objective of cognitive radio is to facilitate an efficient utilization of the wireless spectrum through a highly reliable method [20].

Cognitive radio is the future wireless systems (device) that autonomously monitor the spectrum; it recognizes the spectrum which is not used by the legacy (primary) user and uses this unused band in intelligent way when there is a need. It has 2 primary objectives:

(i) Highly reliable communication at anywhere and anytime needed.

(ii) Effective use of the spectrum.

An important concept is the definition of spectrum hole. It can be defined as: 'a spectrum hole is a set of frequencies allocated to a primary user (licensed user), but, in specified time and in specific geographic location, the frequency is not being used by that user'.

The efficient use of the spectrum will be endorsed by exploiting the spectrum holes. At particular instant if primary user wants the spectrum hole then cognitive user has to move to another spectrum hole or stay in the same band while changing its transmission parameters to avoid the interference with legacy user. Spectrum hole can also be called as this process is illustrated in figure1 [20].

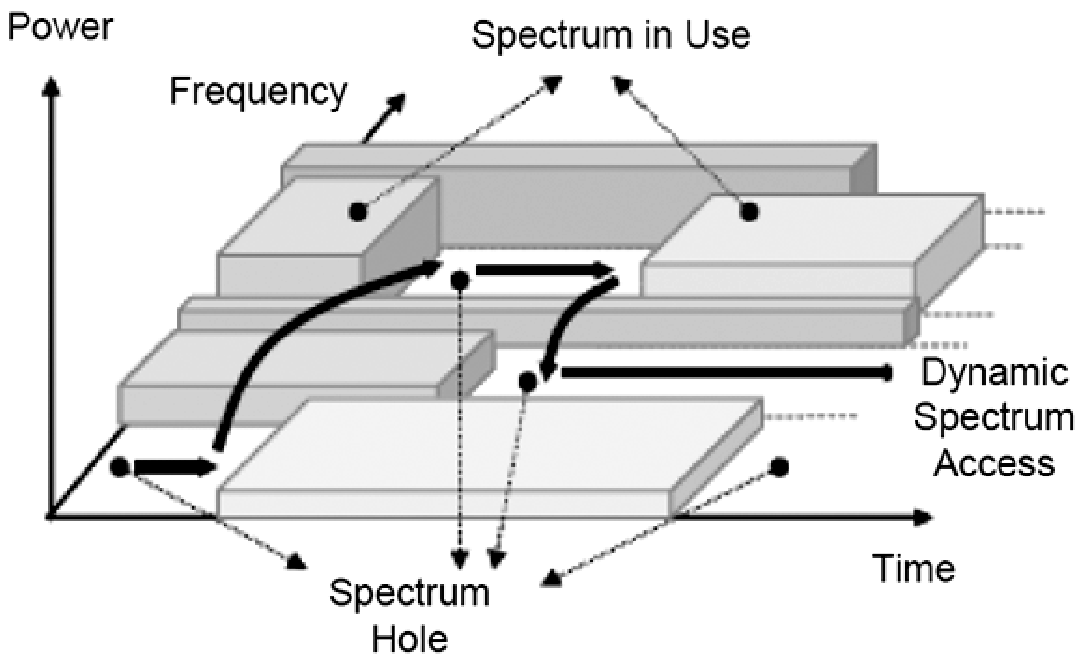

Figure 1: Example of the Utilization of Spectrum holes

A Dynamic Spectrum Access (DSA) network, fortified with a Cognition Cycle (CC) formulate a CR network; hence, the basic and main idea of cognitive radios is based on the cognitive cycle, according to which radios must be able to observe their operating environment, then decide how best to adapt to it, and act consequently. As the cycle repeats, the radio should be able to learn from its previous actions. The principle rests on the radio's ability to observe, adapt, reason, and learn.

The CC achieves context awareness such that each SU can provide spectrum sensing and recognizes the unused spectrum (white space) in its operating environment and intelligence so that each SU is capable of using the high quality white space in order to get maximum the SU network performance, and to minimum interference to the Pus [2].

Study of Cognitive engine helps us to better understand the functionality of Cognitive radio. The cognitive engine works according to the principle of cognitive cycle. It consists of several steps like analyzing the radio frequency stimulate from the other environment and detecting the spectrum holes. It also includes functions like transmission power control and managing the spectrum so that efficient spaces are utilized and allocated to the unlicensed users and also to ensure interference free opportunistic spectrum access [11].

Figure 2 illustrates the functioning of cognitive cycle. The cognitive cycle consists of the following spectrum functions: -

1. Spectrum sensing: Cognitive radio senses the whole spectrum and find out 'spectrum holes' (frequency bands not used by licensed users).

2. Spectrum analysis: Based on the information provided by spectrum sensing, cognitive radio will guess the channel state and the channel capacity.

3. Spectrum decision: The decision-making part is the core research area in CR. According to the previous information provided by spectrum sensing and spectrum analysis, cognitive radio needs to conclude not only which available channel to use but also the transmission parameters, e.g. the transmission mode, the data rate and transmission power etc.

After the above 3 steps, cognitive radio will have enough information to adjust its operating parameters to perform the communication. The cognition part is the intelligence concentrated part of cognitive radio where different intelligent techniques are applied, including reasoning and learning. The decisions made by individual users will change the environment and other users will adjust themselves to these changes by going through the 3 steps frequently.

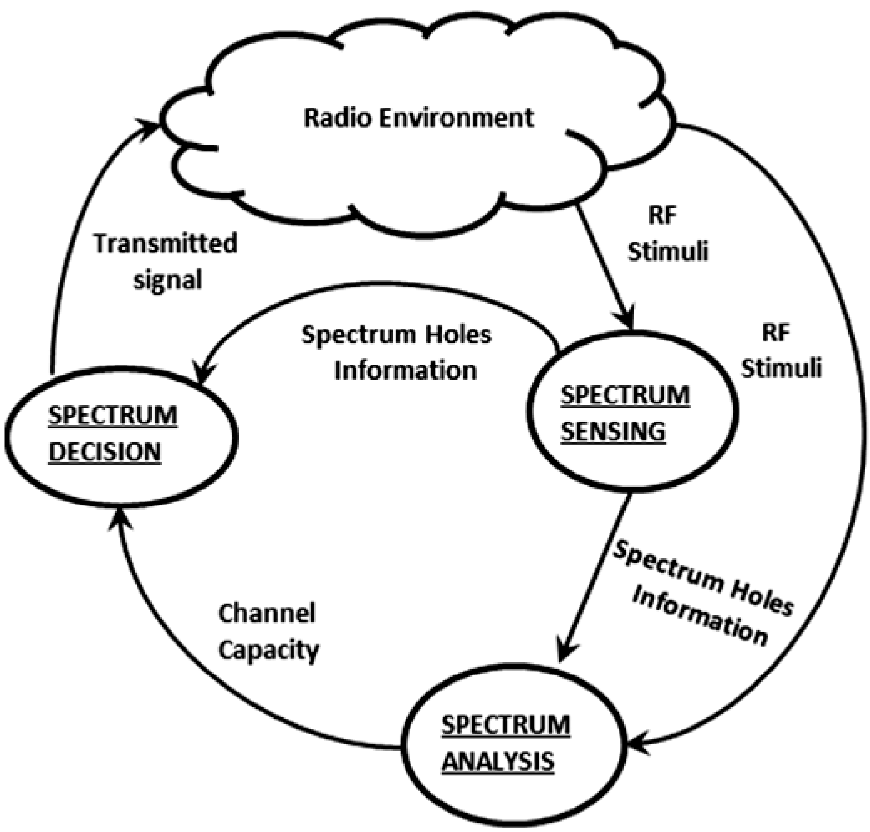

Figure 2: Basic Cognitive Cycles

2.1. CRN Architecture

Cognitive Radio Network architecture helps to improve the use of the spectrum and also helps to boost-up the performance of the network. From user point of view, the network utilization means that they can always satisfy their demands anytime and anywhere through accessing CRN. From the service provider view, they can not only provide better services to their users (customers). The basic component of CRN architecture is base station, mobile station and backbone networks. There are three basic types of network architecture in CRN are shown below [11].

2.1.1. Infrastructure Architecture

In this type of architecture, a MS can directly access a BS/AP in the one hop manner. MSs under the transmission range of the same BS/AP shall communicate with each other through the BS/AP.

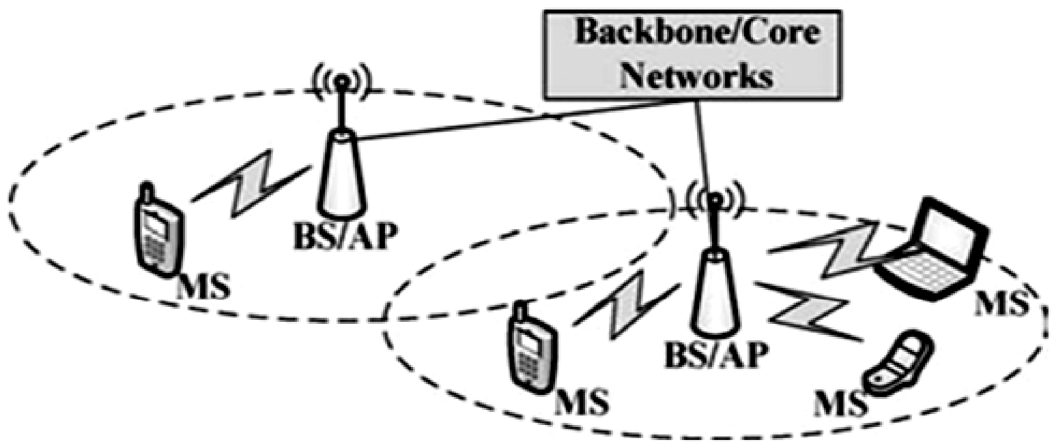

Figure 3: Infrastructure architecture of a CRN

Communications between different cells are routed through backbone/core network. The BS/AP may be able to run one or multiple communication standards/protocols to satisfy different demands from MSs. A CR terminal can also access various kinds of communication through their respective BS/AP and this type of architecture is shown in Figure 3.

2.1.2. Ad-hoc Architecture

There will not be any infrastructure support in ad-hoc architecture (Figure 4). The network is set up on the fly. If a MS identifies that there are some other MSs nearby and they are get connected through certain communication standards/protocols, they can set up a link and as a result ad-hoc network is formed. Note that these links between nodes can be set up by different communication technologies. In addition, two cognitive radio terminals can communicate with each other either by using existing communication protocols (e.g., Wi-Fi, Bluetooth) or dynamically using spectrum holes and this type of architecture is shown in Figure 4.

2.1.3. Mesh Architecture

This type of architecture can be regarded as combo of ad hoc and infrastructure type of architecture which maintains

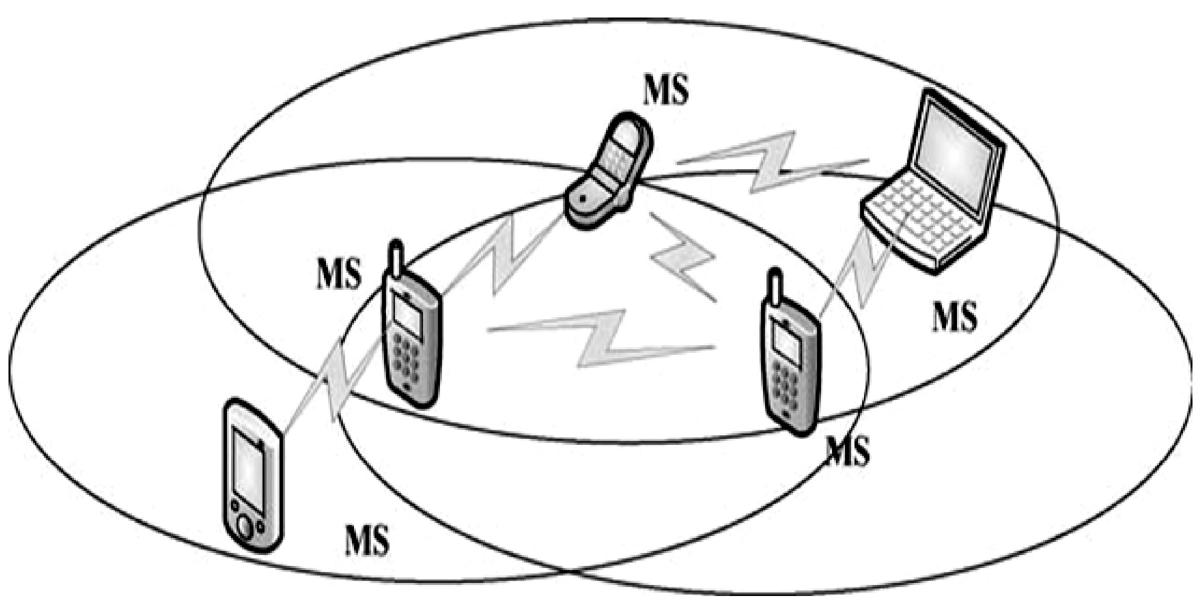

Figure 4: Ad-hoc architecture of a CRN

Connections between the access points, in this type of architecture access point's acts as backbone and behave as and function like wireless routers. In order to access the base station mobile stations, use direct access or use multi hop relay nodes. Some of the base stations which area attached to the wired backbone act as gateways. They behave flexible even without being connected to wire backbone. This architecture offers less cost in establishing the locations of these access points. The combination of cognitive radio capabilities with the base points mainly use holes of the spectrum for the communication purpose. Because of the large availability of the spectrum holes we can be able to serve link between cognitive radio base stations as wireless bones. It comprises both the advantages and disadvantages of the both infrastructure and ad-hoc architecture and this type of architecture is shown in Figure 5.

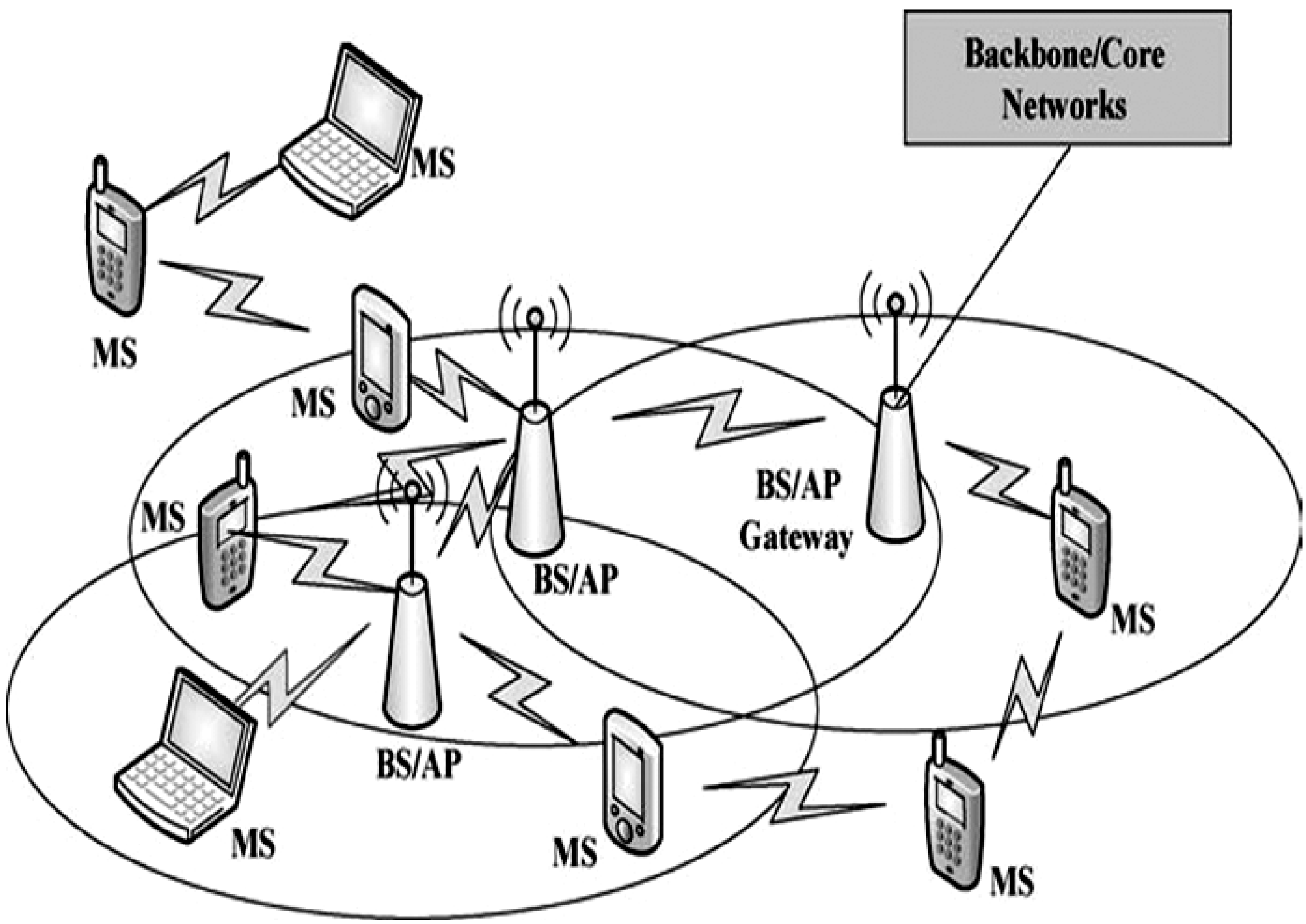

Figure 5: Mesh Architecture

3. REINFORCEMENT LEARNING

The Cognitive Engine is the intellectual system behind the Cognitive Radio, it combines sensing, learning and optimization algorithms to control and adapt the radio system. For this effect, many different learning methods are available and can be used by a Cognitive Radio. They are soft computing techniques, which include: Artificial Neural Networks, evolutionary/Genetic Algorithms, reinforcement learning, fuzzy systems, Hidden Markov Models etc.

RL is an unsupervised machine-learning technique that improves system performance with simple modeling. Here 'unsupervised' means there is no external teacher or critic to direct the learning process. The agent learns about the operating environment by itself [5].

RL has been applied as an alternative to the traditional policy-based methodology for system performance enhancement. The policy-based method requires an agent to follow a set of static and predefined rules on action selection and it has been traditionally applied to a wide range of application schemes in wireless networks.

The vital feature of RL that can enable full self-organization and high adaptability in cognitive wireless networks hypothetically and is its lack of need for any a priori knowledge of the environment model, this feature is the main reason for the widespread use of RL for DSA in wireless networks, since building an accurate analytical model of an arbitrary wireless communications environment is often unfeasible or even impossible [19].

Figure 6 shows a simplified version of a RL model. At instant of time, a learning agent or a decision maker observes state and reward from its operating environment, learns, decides, and carries out the optimal actions. The important representations in the RL model for an agent are as follows.

(i) State represents the decision-making factors, which will affect the reward (or network performance), observed by an agent from the operating environment. Examples of states are the channel utilization level by PUs and channel quality.

(ii) Action represents an agent's action, which may alter or affect the state (or operating environment) and reward (or network performance) and hence the agent learns to take optimal actions at most of the times.

(iii) Reward represents the positive or negative effects of an agent's action on its operating environment in the previous time instant. In other words, it is the result of the previous action on the operating environment in the form of network performance (e.g., throughput).

At any time, instant, an agent observes its state and carries out an appropriate action so that the state and reward, which are the results of the action and will improve in the next time instant. Generally speaking, RL

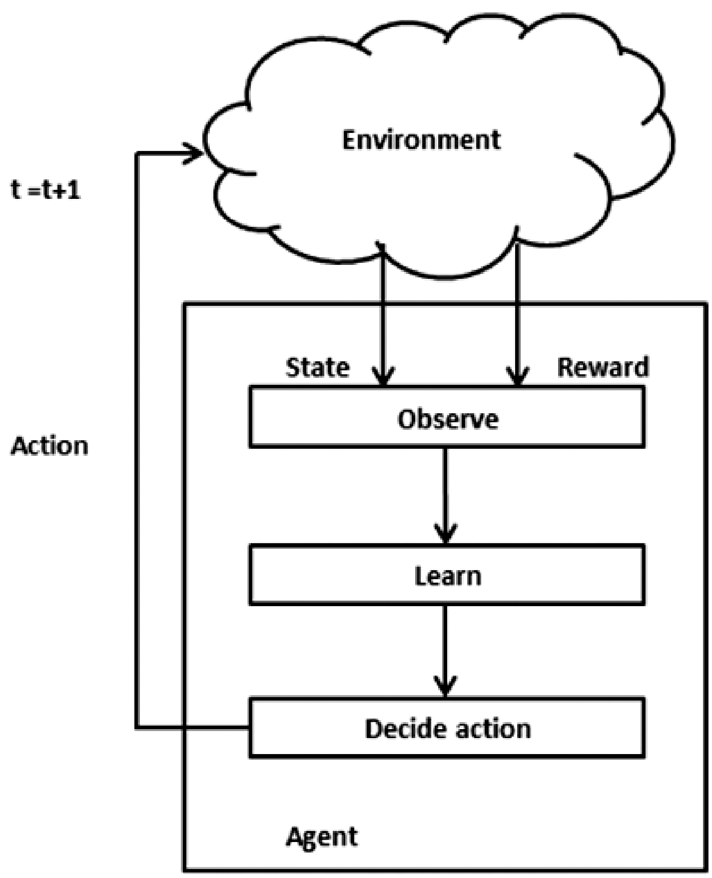

Figure 6: simplified RL model

approximates the reward of each state-action pair, and this formulates knowledge. The most important component in Figure 1 is the learning engine that provides knowledge to the agent. [1] Briefly describes how an agent learns.

There are two RL methods such as, model based and model-free, which perform optimization in very different ways.

Model-based RL is capable of constructing an internal model of the transitions and immediate outcomes in the environment with the help of experience. Suitable actions are then chosen by searching or planning in this world model. This is a statistically efficient way to use experience, as each piece of information from the environment can be stored in a statistically faithful and computationally operable way. Provided that constant re-planning is possible, this allows action selection to be readily adaptive to changes in the transition contingencies and the utilities of the outcomes. This flexibility makes model-based RL suitable for supporting goal-directed actions. For instance, in model-based RL, performance of actions leading to rewards whose utilities have decreased is immediately diminished. Via this identification and other findings, the behavioral neuroscience of such goal directed actions suggests a key role in model-based RL.

Model-free RL, on the other hand, uses experience to learn directly one or both of two simpler quantities (state/ action values or policies) which can achieve the same optimal behavior but without estimation or use of a world model. Given a policy, a state has a value, defined in terms of the future utility that is expected to accrue starting from that state. Crucially, correct values satisfy a set of mutual consistency conditions: a state can have a high value only if the actions specified by the policy at that state lead to immediate outcomes with high utilities, and/or states which promise large future expected utilities have high values themselves). Model-free learning rules such as the temporal difference (TD) rule define any momentary inconsistency as a prediction error, and use it to specify rules for plasticity that allow learning of more accurate values and decreased inconsistencies. Given correct values, it is possible to improve the policy by preferring those actions that lead to higher utility outcomes and higher valued states. Direct model-free methods for improving policies without even acquiring values are also known.

The advantages of RL are as follows:

(i) It does not require previous knowledge of the operating environment so a SU can learn the operating environment by itself.

(ii) RL modeling is very simple compare to other supervised learning technique.

(iii) RL has inherent cognitive capabilities since it can learn from the suitable set of actions in order to maximize a numerical reward. So RL has been applied to spectrum sensing or spectrum sharing procedures in Cognitive Radio (CR) instead of other learning technique [11].

The disadvantages of RL are as follows:

(i) The system using reinforcement learning take a Longer learning times to provide credible results than simple learning.

3.1. Performance Enhancements

The RL approach has the following performance enhancement.

(P1) Throughput or Good put: Will get higher throughput (or good put) only when packet delivery rate and successful packet transmission rate are higher and packet loss rate should be minimum, reduce the convergence time to get higher throughput [6, 16].

*(P2) End-to-End Delay/Link Delay: E*nd-to end delay, which is the total link delays along a route, indicates shorter time duration for packets to travel from a source node to its destination node which must be of lower value [11(By using CTBR protocol)].

(P3) Interference to Pus: The level of interference to PUs must be low such that the number of collisions with PU activities are low [15(By using HMM method)].

(P4) Sensing Channels: The number of sensing channels

Should be low indicates lower sensing overheads (i.e., delays and energy consumption) [7].

(P5) Overall Spectrum Utilization: In order to increase the overall spectrum utilization, increase channel access time, while reduce blocking and dropping of calls, respectively [2, 3, 5].

*(P6) Number of Channel Switches: R*educe number of channel switches in order to reduce channel switching time.

(P7) Energy Consumption: Lower energy consumption indicates longer network lifetime and number of survival nodes.

(P8) False Alarm: Reduce false alarm, which occurs when a PU is mistakenly present in an available channel, in channel sensing (A2).

(P9) PU Detection: Increase the probability of PU detection in order to reduce miss detection in channel sensing (A2). Miss detection occurs whenever a PU is mistakenly considered absent in a channel with PU activities [16,18].

(P10) Number of Channels Being Sensed Idle: Increase the number of channels being sensed idle, which contains more white spaces [2, 5].

(P11) Accumulated Rewards: Increase the accumulated rewards, which represent gains, such as throughput performance. Improve SU's total payoff, which is the difference between gained rewards (or revenue) and total cost incurred [3, 10].

3.2. Application schemes of RL to Cognitive Radio Network

(A1) Dynamic Channel Selection (DCS): The DCS scheme selects operating channel(s) with white spaces for data transmission whilst minimizing interference to PUs. The [2, 5, 8] propose a DCS scheme that enables SUs to learn and select channels with low packet error rate and lowest level of channel utilization by PUs in order to enhance QoS, particularly throughput and delay performances.

(A2) Channel Sensing: Channel sensing senses for white spaces and it also detect the presence of PU activities. In [7], the SU reduces the number of sensing channels and may even turn off channel sensing function if its operating channel has achieved the required successful transmission rate in order to enhance throughput performance. The SU determines the durations of channel sensing, time of channel switching, and data transmission, respectively, in order to enhance QoS, particularly throughput, delay, and packet delivery rate performances. Both incorporate DCS (A1) into channel sensing in order to select operating channels. Due to the environmental factors that can deteriorate transmissions (e.g., multipath fading and shadowing), [7] propose a cooperative channel sensing scheme, which combines sensing results from cooperating one hop SUs, to enhance the accuracy of PU detection.

(A3) Security Enhancement: Security enhancement scheme [14] aims to enhance the effects of attacks from fake SUs. [14] Propose a security enhancement scheme to minimize the inaccurate sensing outcomes received from neighboring SUs in channel sensing (A2). A SU becomes malicious whenever it sends inaccurate sensing outcomes, intentionally (e.g., Byzantine attacks) or unintentionally (e.g., unreliable devices). [13] propose an anti-jamming scheme to minimize the effects of jamming attacks from malicious SUs, which constantly transmit packets to keep the channels busy at all times so that SUs are deprived of any opportunities to transmit.

(A4) Energy Efficiency Enhancement: Energy efficiency enhancement scheme aims to minimize energy consumption. An energy efficient channel sensing scheme to minimize energy consumption in channel sensing. Energy consumption varies with activities, and it increases from sleep, idle, to channel sensing. The scheme takes into account the PU and SU traffic patterns and determines whether a SU should enter sleep, idle, or channel sensing modes. Switching between modes should be minimized because each transition between modes incurs time delays.

(A5) Channel Auction: Channel auction provides a bidding platform for SUs to compete for white spaces. The channel auction scheme that enables the SUs to learn the policy (or action selection) of their respective SU competitors and place bids for white spaces. This helps to allocate white spaces among the SUs efficiently and fairly.

(A6) Medium Access Control (MAC): MAC protocol aims to minimize packet collision and maximize channel utilization in CR networks. A collision reduction scheme that reduces the probability of packet collision among PUs and SUs, and it also increase throughput and decrease packet loss rate among the SUs. [4] Proposes a retransmission policy that enables a SU to determine how long it should wait before transmission in order to minimize channel contention.

(A7) Routing: Routing enables each SU source or intermediate node to select its next hop for transmission in order to search for the best route(s), which normally incurs the least cost or provides the highest amount of rewards, to the SU destination node, Each link within a route has different types and levels of costs, such as queuing delay, available bandwidth or congestion level, packet loss rate, energy consumption level, and link reliability, as well as changes in network topology as a result of irregular node's movement speed and direction [11, 17].

(A8) Power Control: The power selection scheme that selects an available channel and a power level for data transmission, the purpose is to improve its Signal-to-Noise Ratio (SNR) that results in improved packet delivery rate.

4. INFERENCES

Paper	*Methodology*	*Advantages and disadvantages and disadvantages*	*Future scope*
	Basics of Reinforcement Learning ,features ,performance enhancements and application schemes		
[1]	RL model, components, features, different application schemes, performance enhancements has been discussed.	Advantages: 1. Higher Throughput/Good put. 2. End-to-End Delay/Link Delay must be low. 3. Low Level of Interference to Pus. 4. Lower Number of Sensing Channels. 5. Higher Overall Spectrum Utilization. 6. Lower Number of Channel Switches. 7. Lower Energy Consumption. 8. Lower Probability of False Alarm. 9. Higher Accumulated Rewards.	1. The two-layered multi-agent RL model can be used in CR network applications. 2. Apply or integrate the RL features and enhancements (e.g., state, action, and reward representat-ions) to other learning-based approaches, such as the neural network-based approach. 3. The RL models and algorithms applied to other kinds of networks such as cellular radio access networks and sensor networks, which may able to provide performance enhancement in CR networks.
[2]	How RL is used to implement the cognitive cycle in CRN is discussed.	Advantages: 1. RL adopts a simple modeling	1. It can be used in the cross layer

(*contd...Table 1*)

Paper	*Methodology*	*Advantages and disadvantages and disadvantages*	*Future scope*
		approach. Thus the complexity involved in modeling is minimized. 2. RL makes an agent to adapt to its uncertain operating environment.	application in CR networks such as DCS, topology management, and congestion control and scheduling.
[4]	Developing an enhanced version of actor critic learning algorithm called as continues actor critic learning and comparing this with traditional ACL and Q learning scheme.	Advantages: 1. CACLA scheme can achieve high throughput performance and less execution time compared to other two schemes. *Disadvantages*: 1. CACLA reduces variance of the observed performance parameters.	
		Prediction of spectrum sensing	
[15]	Prediction in cognitive radio networks is a challenging task that involves several subtopics such as channel status prediction, PU activity prediction, radio environment prediction and transmission rate prediction. We present an overview on the most important prediction techniques in CRN.	Disadvantages: 1. Because of heterogeneous property of the CR users and the uncertainty property of the CR communications, it’s difficult to predict the service requests of the CR users in time, space, and frequency domains. Thus, it is difficult to coordinate the spectrum sharing between CR users through prediction.	
		Spectrum sensing	
[7]	RL based co-operative sensing method is to state co-operation overhead problem and improves the co-operative gain in CR network.	Advantages: 1. RLCS method reduces the overhead of cooperative sensing while effectively improving the detection performance. 2. If there is any change in environment then also it can capable of adapting to that environment and maintains comparable performance under the influence of primary user activity, user movement, user reliability and control channel fading.	
		DCS, OSA, DSA, Spectrum assignment, Spectrum sharing	
[3]	It introduces a distributed spectrum sharing scheme in the context of CR which enables effective use of spectrum and which is achieved using RL.	Advantages: 1. Spectrum sharing scheme can able to reduce the need for spectrum sensing which in turn save the power and time for sensing.	
[5]	It states that how RL is used to achieve context awareness and intelligence in wireless network and also states how the problems (resource management, routing, DCS) in wireless network are solved using RL is discussed.	Advantages: 1. RL is a simple since it is model free approach. 2. Complexity is reduced.	Reinforcement learning can be applied in wireless networks including mobile ad-hoc networks (MANETs), wireless sensor networks (WSNs), cellular networks, and recently the next generation wireless networks, such as cognitive radio networks.

(*contd...Table 1*)

Paper	*Methodology*	*Advantages and disadvantages and disadvantages*	*Future scope*
[6]	It illustrates a spectrum sharing protocol for secondary users combining multichannel Non-persistent carrier sense multiple access (CSMA) and reinforcement learning.	Advantages: The Reinforcement Learning conjunction with non-persistent CSMA will improve the throughput.	
[8]	It describes the use of RL application schemes that has necessity of context-awareness and intelligence such as the Dynamic Channel Selection (DCS), scheduling, and congestion control.	Advantages: 1. Our enhanced RL approach provide network wide performance enhancement in throughput and stability with significant reduced number of channel switching's.	
[9]	Distributed framework for spectrum assignment in the context of cellular primary networks has been presented.		
[10]	It uses the repeated game modeling multiuser dynamic spectrum accessing and proposes a multi agent reinforcement learning method (multiuser independent Q-learning method) with which the CR user coordinates in choosing best highest gain channel and avoiding conflict between each other.	Advantages: 1. The user action can converge to Nash Equilibrium with high probability and achieved total reward is close to the maximal reward with proposed MIQ algorithm.	
[12]	It presents a machine learning-based scheme that wills advent the cognitive radios capabilities to enable effective OSA, so it improves the efficiency of spectrum utilization.	Advantages: The proposed learning technique does not require previous knowledge of the environment's characteristics and dynamics, still also it can achieve high performance by learning from interaction with the environment	
[18]	We address the decision making criteria for a secondary user (SU) for deciding when to transmit or not depend upon performing spectrum sensing and identifying the presence of any primary users in the environment in a cognitive radio network.	Advantages: RL significantly improves the performance of SU transmissions by considering the probability of interference and wastage of spectrum.	
		Jamming and their mitigation	
[13]	We propose a strategy to avoid or mitigate reactive forms of jamming using RL.	Advantages: 1. By using a learning approach, there is no need to pre-program radio with specific anti-jam strategies and the problem of having to classify jammers is avoided.	This strategy is well-suited for frequency hopping spread spectrum systems.
		Routing	
[11]	Presented some routing protocols in CRN based on their classification and operation.		
[17]	Presented some recent routing protocols in CRN		

5. CONCLUSION

Reinforcement learning (RL) has been applied in Cognitive radio (CR) networks to achieve context awareness and intelligence. Examples of schemes are dynamic channel selection, channel sensing, security enhancement mechanism, energy efficiency enhancement mechanism, channel auction mechanism, medium access control, routing, and power control mechanism. To apply the RL methods, several illustrations may be necessary including state and action, as well as rewards. Based on the CR context, this paper presents an extensive review on the performance enhancements.

REFERENCES

[1] Kok-Lim Alvin Yau,1 Geong-Sen Poh,2 Su Fong Chien,3 and Hasan A. A. Al-Rawi1 "Application of Reinforcement Learning in Cognitive Radio Networks: Models and Algorithms" Hindawi Publishing Corporation ?e Scientific World Journal Volume 2014.

[2] Kok-Lim Alvin Yau, Peter Komisarczuk and Paul D. Teal "Applications of reinforcement learning to Cognitive Radio Networks". *Workshop on Cognitive Radio Interfaces and Signal Processing (CRISP) at the International Conference on Communications (ICC'10) IEEE*, Cape Town, South Africa, May 2010.

[3] Tao Jiang, David Grace, Yiming Liu "Performance of Cognitive Radio Reinforcement Spectrum Sharing Using Different Weighting Factors" International Conference on Communications and Networking in China, 2008. ChinaCom 2008.Third Conference DOI: 10.1109/CHINACOM.2008.4685240

[4] Dhananjay Kumar1, Pavithra Hari2, Panbhazhagi Selvaraj3 and Sharavanti Baskaran "Efficient spectrum utilization in cognitive radio through reinforcement learning" ICTACT international journal on communication technology, september 2013, volume: 04, issue: 03.

[5] Kok-Lim AlvinYau a, PeterKomisarczuk a,b, PaulD.Teal a,n "Reinforcement learning for context awareness and intelligence in wireless networks: Review, new features and open issues" 25th International Conference on Advanced Information Networking and Applications (AINA'11) *IEEE*, Biopolis, Singapore, March 2011.

[6] Nisha Kiran1, Ravi Raj2, Alisha Khan3 "Maximization of Throughput in Cognitive radio using non-persistent CSMA combined with Reinforcement Learning" International journal of innovative research in electrical, electronics, instrumentation and control engineering vol. 3, issue 5, may 2015.

[7] Brandon F. Lo • Ian F. Akyildiz "Reinforcement learning for Cooperative sensing gain in Cognitive Radio ad hoc networks" Springer Science+Business Media New York 2012.

[8] Kok-Lim Alvin Yau, Peter Komisarczuk and Paul D. Teal "Context-Awareness and ntelligence in Distributed Cognitive Radio Networks: A Reinforcement Learning Approach" 11th Australian Communications Theory Workshop (AusCTW'10) IEEE,Canberra, Australia, February 2010.

[9] Francisco Bernardo, Ramon Agust´ý, Jordi P´erez-Romero and Oriol Sallent "Distributed Spectrum Management based on Reinforcement Learning" Proceedings of The 4th International Conference On Crowncom 2009

[10] 1 Wu Chun, 2 Yin Mingyong, 2 Ma Shaoliang, 1 Jiang Hong "Multiagent Reinforcement Learning Dynamic Spectrum Access in Cognitive Radios" Sensors & Transducers, Vol.164, Issue 2, February 2014, pp. 170-175.

[11] T. Pavan Kumar#1 E. Suresh babu#2 B. Venkata Ramana#3B. Sai Shashank #4 "Survey: Routing Protocols in Cognitive Radio Mesh Networks" T. Pavan Kumar et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (1), 2015, 603-608.

[12] Pavithra Venkatraman, Bechir Hamdaoui, and Mohsen Guizani "Opportunistic Bandwidth Sharing Through Reinforcement Learning".

[13] Marc Lichtman and Jeffrey H. Reed "Reinforcement Learning for Reactive Jamming Mitigation" Journal of Cyber Security, Vol. 3 No. 2, 213–230 @2014 River Publishers.

[14] Mee Hong Linga, ", Kok-Lim Alvin Yaua, Junaid Qadirb, Geong Sen Pohc, Qiang Nid "Application of reinforcement learning for security enhancement in cognitive radio Networks" M.H. Ling et al. / Applied Soft Computing 37 (2015) 809–829 811

[15] Xiaoshuang Xing1, Tao Jing1, Wei Cheng2 Yan Huo1, Xiuzhen Cheng3 "Spectrum prediction in Cognitive Radio Networks"

[16] Yu Ren, Pawel Dmochowski, Peter Komisarczuk "Analysis and Implementation of Reinforcement Learning on a GNU Radio Cognitive Radio Platform" CROWNCOM2010.9170 Digital Object Identifier: 10.4108/ICST

[17] Apurva Varade, Shriram Kulkarni "overview and literature survey on routing Protocols for cognitive radio networks"

International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163 Volume 1 Issue 10 (November 2014)

[18] Senthuran Arunthavanathan, y Sithamparanathan Kandeepan, m& Robin J. Evans “Reinforcement Learning based Secondary User Transmissions in Cognitive Radio Networks” 2013 IEEE GLOBECOM workshops (gc wkshps)

[19] Thesis by Nils Morozs Ph.D. University of York Electronics September 2015 “Accelerating Reinforcement Learning for Dynamic Spectrum Access in CognitiveWireless Networks”.

[20] Thesis by Tao Jiang Doctor of Philosophy (Ph.D.) University of York Department of Electronics September 2011 “Reinforcement Learning-based Spectrum Sharing for Cognitive Radio.

[21] “Cognitive Radio Networks” by Kwang –Cheng Chen and Ramjee Prasad.