Open Access **Baghdad Science Journal** **P-ISSN: 2078-8665**
2021, Vol. 18 No.2 (Suppl. June) **E-ISSN: 2411-7986**

DOI: http://dx.doi.org/10.21123/bsj.2021.18.2(Suppl.).0947

Reinforcement Learning-Based Television White Space Database

***Armie E. Pakzad*[1*]** ***Raine Mattheus Manuel*[1]** ***Jerrick Spencer Uy*[1]**
***Xavier Francis Asuncion*[1]** ***Joshua Vincent Ligayo*[1]** ***Lawrence Materum*[1,2]**

[1] De La Salle University, Philippines.
[2] Tokyo City University, Japan.
[*]Corresponding author: armie.pakzad@dlsu.edu.ph, raine_manuel@dlsu.edu.ph, jerrick_uy@dlsu.edu.ph, francis_asuncion@dlsu.edu.ph, joshua_ligayo@dlsu.edu.ph, materuml@dlsu.edu.ph
[*]ORCID ID: https://orcid.org/0000-0002-6230-9092[*], https://orcid.org/0000-0002-6039-8642, https://orcid.org/0000-0002-5815-5414, https://orcid.org/0000-0003-0923-0622, https://orcid.org/0000-0003-2034-3863, https://orcid.org/0000-0001-6832-2610

Received 28/3/2021, Accepted 13/4/2021, Published 20/6/2021

Abstract:

Television white spaces (TVWSs) refer to the unused part of the spectrum under the very high frequency (VHF) and ultra-high frequency (UHF) bands. TVWS are frequencies under licenced primary users (PUs) that are not being used and are available for secondary users (SUs). There are several ways of implementing TVWS in communications, one of which is the use of TVWS database (TVWSDB). The primary purpose of TVWSDB is to protect PUs from interference with SUs. There are several geolocation databases available for this purpose. However, it is unclear if those databases have the prediction feature that gives TVWSDB the capability of decreasing the number of inquiries from SUs. With this in mind, the authors present a reinforcement learning-based TVWSDB. Reinforcement learning (RL) is a machine learning technique that focuses on what has been done based on mapping situations to actions to obtain the highest reward. The learning process was conducted by trying out the actions to gain the reward instead of being told what to do. The actions may directly affect the rewards and future rewards. Based on the results, this algorithm effectively searched the most optimal channel for the SUs in query with the minimum search duration. This paper presents the advantage of using a machine learning approach in TVWSDB with an accurate and faster-searching capability for the available TVWS channels intended for SUs.

Key words: Radio propagation, Radio spectrum management, Reinforcement learning, Television white space database

Introduction:

Spectrum scarcity has long been anticipated and spectrum management mechanisms can lead to less optimal outcomes. The upsurge in mobile communications has led to the exploitation of spectrum bands; and thus, the efficient use of radio spectrum for wireless data communication faces challenges nowadays (1). Efficient spectrum use is the current issue even though more than 70% of the spectrum are not fully utilised and available for other users and services (2). The utilisation of other bands is minimal, especially the very high frequency (VHF) to ultra-high frequency (UHF) bands. Specific channels are left vacant for some time, which can be used for wireless applications. These vacant channels are referred to as television white space (TVWS), which makes a better option for wireless communications since it can penetrate obstructions, such as buildings, terrains, and the likes. Therefore, it can successfully enhance the effective use of spectrum and resolve spectrum scarcity worldwide (3). The use of TVWS communications is seen to improve the delivery of services in rural areas. Only 15% of the TV spectrum are utilised at present. Thus, it could be a great way to provide connectivity to almost 3 billion people across the globe (4).

The rapid growth in wireless communications increases the demand for more efficient spectrum management (5). The migration to digital TV can lead to nearly an additional 300 MHz of spectrum bandwidth. One factor to consider in using vacant TV bands is protecting primary users (PUs) from interference with secondary users (SUs) (6). TVWS is a distinguished new opportunity for wireless

communications due to its capability of penetrating obstacles as well as its low utilisation at given times and specific areas.

One of the promising techniques in implementing TVWS in communications is the use of the TVWS database (TVWSDB). Its main purpose is to protect PUs from interference with SUs. There are several geolocation databases available for this purpose. Nevertheless, it is unclear if those databases have the prediction feature that gives TVWSDB the capability of decreasing the number of inquiries from SUs. With this in mind, the authors of the this paper present a reinforcement learning-based TVWSDB. A comparison between the reinforcement learning-based (RL) and a conventional TVWSDB is carried out to determine the suitable comparative performance metrics to be considered for future work.

This paper is organised as follows: Section 2 presents the literature review; Section 3 discusses the RL technique; Section 4 explains the methodology and the RL algorithm used in this paper; and Section 5 discusses the conclusion and future work.

Literature Review:

This section discusses several studies conducted on the utilisation of TVWS for communications. The following are the significant works related to this study.

The implementation of TVWS is ongoing in some countries while others are still in the trial phase. In Mueck and Noguet (2011), an overview of a European framework related to the efficient exploitation of TVWS was presented, while Anabi et al. (2016) tackled the challenges, trends, and future of database-assisted TVWS technology (7,6). It was discussed that this type of TVWS technology set the standard for an interference-free system for users. It examined the operations of a database-assisted TVWS system, which consisted of two types of TVWS devices (TVWSDs), namely master and slave. The master TVWSD had direct access to the TVWS operator, while the slave TVWSD could only have access to the TVWS connection through the master. Other than the TVWSDs, the system also had a TVWS operator and TVWS database (TVWSDB). The TVWS operator had access to the database while the database contained information on different TVWS bands available and what transmission power to use.

The process in geolocation-assisted TVWS technology began with the master TVWSD inquiring for a list of TVWS operators. The master TVWSD then sent queries for available bands to use to the operator. The results were dependent on TVWSD's provided location and operational parameters. The TVWS operator then had to access a TVWSDB where it was provided with available bands and transmission power to use. The operator would make necessary calculations to determine which frequency and transmission power to use for the master TVWSD. After determining the parameters for the master TVWSD, a slave TVWSD would query the master regarding what frequency and transmission parameters to obtain connectivity. After giving information to the master TVWSD, the master provided operational parameters to the slave device, allowing the slave device to finally have connectivity (6).

The Federal Communications Commission (FCC) required TV band devices to access unutilised TV channels through a geolocation database (GDB). The FCC rule of adopting the database architecture brings challenges in designing and implementing TVWS technology. In a study conducted by Feng et al. (8), the researchers presented WhiteNet. This system was compatible with the architecture of the FCC database. It was stated that WhiteNet was a TVWS network-accessible system comprised of multiple access points. WhiteNet included a WhiteNet Local Database consisting of three databases: Vacant TV Channel Database, Local AP Database, and Contention Database.

The WhiteNet Local Database could be integrated into the FCC's database architecture and could resolve spectrum overlapping among the access points. According to the researchers, the WhiteNet Local Database added information on the access points to the standard GDB. The information included the location of an access point, interference between access points, and the channels used by access points. Furthermore, WhiteNet utilised B-SAFE, a distributed spectrum allocation algorithm, to allow access points to determine their spectrum. After implementing WhiteNet, it was concluded that WhiteNet offered a novel access point discovery and selection process, which allowed users to identify multiple access points and provide which access point had the best performance. In GDBs, a device stores information on the availability of TVWS in the form of location, transmission power levels, frequency range, and time. The device can transmit information at a specific frequency range without interference from other channels using this information.

The study by Sun et al. (2012) presented a system architecture and procedure for managing numerous SU networks (9). The study proposed a coexistence management scheme for SUs to share the spectrum opportunity and coexistence among

PUs and SUs. The proposed coexistence scheme was an interference-constrained time allocation scheme. In the study, the researchers compared their proposed time slot allocation scheme to three other schemes through simulations. A new parameter was introduced, known as Quality of Coexistence (QoC), which was used to measure rewards achieved by a scheme. The simulation results illustrated that their proposed coexistence scheme maximised the QoC by dividing a time slot into two time slots and providing them to two SU networks for allocation.

The main reason GDB is used in TVWS is to ensure that white space devices (WSDs) do not interfere with used frequency bands. To avoid interference, a protocol is followed that can be classified into three zones: Exclusion Zone, Restriction Zone, and Protection Zone. In the exclusion zone, no transmission by incumbents should be made. In the restriction zone, specific operating parameters like transmission power, antenna height, and frequency range should be followed. In the protection zone, a maximum level of interference should be obeyed (10).

A study by Ojaniemi et al. (2012) designed a method to raise the precision of GDBs through simulation (11). This method was conducted by creating a database from WSD measurements. By increasing the number of samples, the propagation map could be approximated. Their study was able to showcase that using an algorithm to update the GDB through sensing ensured that PUs and SUs did not suffer from interference.

Open access to TVWS has its advantages, especially in providing wireless communications. The challenges of having an open access feature are security and secondary coexistence. To improve spectrum utilisation, a technology called device-to-device (D2D) communication can be used. A research paper by Xue and Wang (2015) studied the resource sharing problem for many D2D communications by using the TVWS GDB (12). The researchers proposed a system model, together with resource allocation and admission control scheme. These algorithms (evolutionary computation) assisted in spectrum assignment with minimisation of interference and efficient power and admission control. The main goal was to maximise the number of D2D links by properly allocating available channels and coordinating the power levels. The researchers also took note of the quality of the service requirements for the users. Their research was tested in a simulation that showed how well their proposed schemes were.

Contrarywise, Puspita et al. (2019) conducted a study on cognitive radio that involved the use of RL algorithm (13). Their research presented insight on an RL-based cognitive radio network system to provide the most efficient frequency to use for the most efficient spectrum management. They used the deep RL approach to form deep Q-networks. They also showed different survey papers on using artificial intelligence (AI) techniques for cognitive radio.

Multiple studies have already been made with the implementation of computational intelligence in their research. An example would be using sensing algorithms to assure that SUs would not interfere with PUs. Sensing algorithms, which are under the category of neural networks in computational intelligence, were tied with the utilisation of GDBs to continually update the list of vacant frequencies that could be used by SUs (14). A study conducted by Martin et al. (2013) focused on reducing interference between PUs and Sus (15). The only difference between their research and Dionísio et al.'s was that they utilised a fuzzy logic model that produced an enhanced detection algorithm.

Their results revealed that their new algorithm yielded much better results than the traditional PU detection algorithms. Another study involved fuzzy logic-tackled spectrum mobility, or the spectrum handoff, which managed the SUs occupying the available spectrum without interfering with the regular transmission of the PUs. Under spectrum mobility were two Fuzzy Logic Controllers (FLC): price negotiation and duration negotiation. Price and duration were the main input parameters for band-sharing negotiations. The FLC consisted of three main stages: fuzzification, knowledge-based interference, and defuzzification. The operation of the first FLC was to calculate the success rate negotiation among the PU and SUs. Whereas, the second FLC estimated the success rate of pricing and duration among the PUs and SUs (16).

Spectrum sensing through a dynamic trust model (DTM) can help sustain a healthy ecosystem between unlicenced users or customer premise equipment (CPE) and licenced users or PUs in a wireless regional area network (WRAN). CPEs can fill the spectrum holes without interference from incumbent users but must vacate when PUs reclaim the spectrum. A study by Wang et al. (2017) suggested a trust model wherein CPEs gained trust based on the behaviour in using a vacant spectrum (17). There were three types of trust according to the model presented: Direct, Indirect, and Incentive. The model could evaluate a CPE's trust and adjust its spectrum allocation based on its behaviour.

Table 1. Studies on geolocation database

Sources	Computational Intelligence Used	Programming Language / Tool Used
Alhammadi et al. (2016) (16)	Fuzzy Logic	MATLAB
Dionísio et al. (2012) (14)	Neural Network	LabVIEW from National Instruments
Martin et al. (2013) (15)	Fuzzy Logic (Enhanced Detection Algorithm)	Simulink from MATLAB
Xue and Wang (2015) (12)	Evolutionary Computing	MATLAB
Wang et al. (2017) (17)	Neural Network	Python

Lacking in the Approaches

Spectrum database is an essential component of architecture design for the utilisation of the spectrum, particularly TVWS. Studies on the TVWS database included in the literature review did not state anything about predicting the databases' operations. GDBs included in the preceding studies did not employ prediction on the availability of TVWSDB with a prediction feature intended for government-industry stakes.

Currently, there is not much research on implementing RL in TVWS. The present study sheds light on implementing a new field of AI, specifically RL, in TVWS. The benefit of this is that the agent can explore the environment through trial and error, and the information obtained from said exploration can be used to improve the overall output.

Reinforcement Learning:

Reinforcement learning (RL) is a machine learning technique that concentrates on the actions based on the mapping of situations to actions to obtain the highest reward. The learning process is conducted by trying out the actions to gain the reward instead of being told what to do. The actions may directly affect the rewards and the subsequent (18).

RL features include trial and error search and delayed reward. It operates on a dynamic dataset from the environment. It learns the best actions that can produce the optimum result. The agent learns, discovers, and works with the dynamic environment. It examines the environment condition, and based on its observation, selects the suitable action to be taken. On the other hand, the environment shifts to another condition and creates a reward for that particular action, which the agent is to obtain. The most recent data help the agent determine if the action was useful enough to be repeated, or wrong and should be avoided. The observation-action-reward cycle continues until learning is achieved (19).

A function can be found inside the agent frame in which it accepts observations as inputs and maps them to actions that can be depicted as outputs. The function consists of elements of the control system. The function is referred to as the policy that determines the appropriate action based on observations. When the actions taken are good, the environment generates a reward. Nevertheless, since the environment is dynamic, it may be altered, thus static mapping may not be the best. The RL algorithm updates the policy based on the selected actions, environmental observations, and the gathered rewards. The agent uses the RL algorithms to learn the optimal policy during its interaction with the environment to take the best action in any given situation, i.e. the action that produces the highest reward in time (19).

Methodology:

This section presents the methodology of the study consisting of design consideration and the processes involved in testing the TVWSDB.

Design Consideration

The architecture of the TVWSDB is shown in Figure 1. The information in the database was sourced from PUs data, TVWSD, or SUs. The operations of the database were based on the machine learning algorithm using the RL technique.

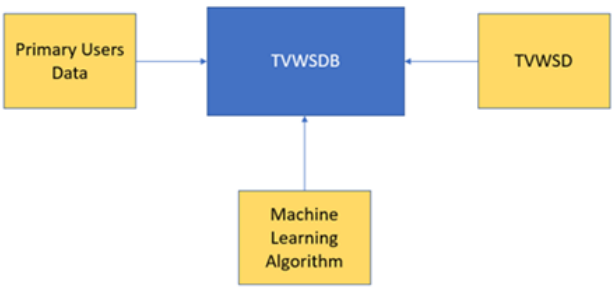

Figure 1. TVWSDB architecture

The TVWSDB

The database stores the data required to determine the available channels for a given location instantly. It contains the parameters of the PUs such as location, frequency, and transmit power, among others. An algorithm processes the data in the TVWSDB to provide the most optimal channel for the inquiring device.

Primary users data

The PUs data provide necessary information regarding the licenced primary users (LPU), which compose the database parameters. For this study, the information came from websites of the LPUs. The contents of the TVWSDB parameters used by the Canadian government (20) were considered

along with those specified in (21). It included data associated with registered TV stations that mainly consisted of the company, channel number, lower and upper frequencies, transmit frequency, location, longitude and latitude, site elevation structure height, transmitter power, and effective radiated power.

The TVWSD (SUs)

The TVWSD is the inquiring device to the TVWSDB, providing necessary information, such as device type, TVWSD ID, TVWSD serial number, TVWSD location, and transmission channel.

The Process

This section discusses the processes done in creating the TVWSDB.

The database in MATLAB

The pertinent data in the TVWSDB were gathered, which included the information that can be seen in Figure 2. The channels that had incomplete information were disregarded as it would be challenging to provide computation. This step narrowed the channels to be used to 13. The availability of each channel was also obtained, which was based on a 00:00 to 23:59 schedule, Monday to Sunday. The data were sourced from different websites and those that were scheduled to be off-air were assumed to be available. If the channel was available, then it was given a numeric value of 1. If the channel was being occupied, it was given a numeric value of 0. Another Excel database was created to host all the necessary information. This database contained all important data needed to identify each channel and the data needed for computation later. The time was split into 30-minute intervals to accommodate the availability from the data in Figure 3. These data were then imported into MATLAB using the readtable function as shown in Figures 4 and 5. The time data imported from the Excel file were converted into a decimal notation; therefore, the datetime function was utilised to convert it back into hour, minute, and second format.

Metro Manila Database

Company	Call Sign	Channel #	Lower Frequency (M	Upper Band (MHz)	Transmit Frequency (MHz)	Location	Latitude	Longitude	Site Elevation (m	Structure Height	Transmitter Power (kw)	ERP
People's Television Network, Inc.	DWGT	4	66.025	72	69.0125	PIA Bldg, Visayas Ave., Q.C.	14.65444	121.04583	56.1	500 ft (150 m)	50kW	500 kW
ABC Development Corp.	DWET	5	76	82	79	Brgy. San Bartolome, Nov. Q.	14.70528	121.04028	62.7	656.168 ft (200m	60kW	120 kW
GMA Network, Inc.	DZBB	7	174.025	180	177.0125	Brgy. Culiat, T. Sora, Q.C.	14.66962	121.05006	56.1	777 ft (236.8 m)	100kW	1000 kW
Radio Philippines Network, Inc.	DZKB	9	186.025	192	189.0125	Panay Ave., Q.C.	14.63833	121.02972	25.6	750 ft (228.6 m)	50kW	10 kW
Zoe Broadcasting Network, Inc.	DZOE	11	198.025	204	201.0125	Brgy. Culiat, Tandang Sora, Q	14.66962	121.05006	59.7	777 ft (236.8 m)	100kW	892.8 kW
Intercontinental Broadcasting Corp	DZTV	13	210.025	216	213.0125	SFDM, Q.C.	14.64944	121.01889	28.5	500 ft (200.5 m)	50kW	500 kW
GMA Network, Inc.	DWDB	27	548.025	554	551.0125	Brgy. Culiat, Q.C.	14.66962	121.05006	59.7	777 ft (236.8 m)	30kW	120 kW
Progressive Broadcasting Corporati	DWAO	37	608.025	614	611.0125	Antipolo City	14.60806	121.16472	218.8	664.37 ft (202.5r	60kW	500 kW

Figure 2. Database information

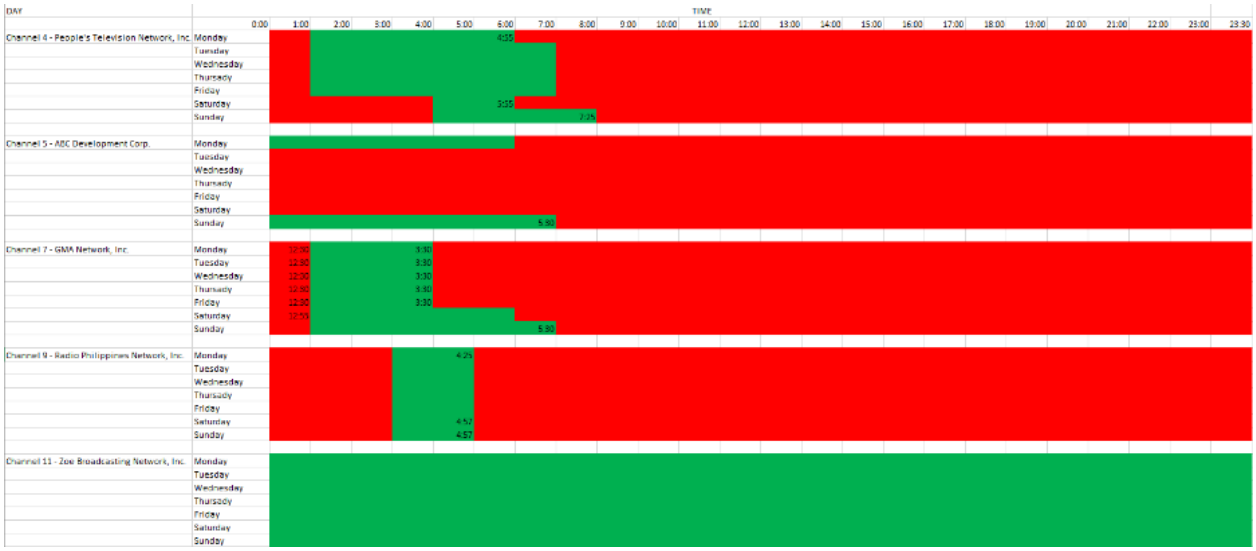

Figure 3. Sample of availability database

DAY	COMPANY	CHANNEL_NUMBER	CALL_SIGN	LOWER_BAND	UPPER_BAND	TRANSMIT_FREQ	LATITUDE	LONGITUDE	ERP	TRANSMIT_POWER	TIME	AVAILABILITY	SECONDARY_AVAILABILITY
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	0:00	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	0:30	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	1:00	1	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	1:30	1	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	2:00	1	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	2:30	1	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	3:00	1	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	3:30	1	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	4:00	1	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	4:30	1	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	5:00	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	5:30	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	6:00	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	6:30	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	7:00	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	7:30	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	8:00	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	8:30	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	9:00	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	9:30	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	10:00	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	10:30	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	11:00	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	11:30	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	12:00	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	12:30	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	13:00	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	13:30	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	14:00	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	14:30	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	15:00	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	15:30	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	16:00	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	16:30	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	17:00	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	17:30	0	0
Monday	People's Television Network, Inc.	4	DWGT	66.025	72	69.0125	14.654444	121.045833	500	50	18:00	0	0

Figure 4. Sample of summarised database

ch4 = 48×15 table

	DAY	COMP...	CHANNEL_NU...	CALL_SI...
1	'Fri'	'People's Tel...	4	'DWGT'
2	'Fri'	'People's Tel...	4	'DWGT'
3	'Fri'	'People's Tel...	4	'DWGT'
4	'Fri'	'People's Tel...	4	'DWGT'
5	'Fri'	'People's Tel...	4	'DWGT'
6	'Fri'	'People's Tel...	4	'DWGT'
7	'Fri'	'People's Tel...	4	'DWGT'
8	'Fri'	'People's Tel...	4	'DWGT'
9	'Fri'	'People's Tel...	4	'DWGT'

Figure 5. Sample of MATLAB database

The RadioPlanner outputs

The RadioPlanner software was used to obtain the range of the transmitters. Information from the created database was used to produce the sample outputs as shown in Figures 6 and 7. The transmitters were created by inputting the name, longitude and latitude, site elevation, frequency, transmitter power, and antenna height. The coverage of the transmitters was given by coloured circular lines in each figure. The data exported from the RadioPlanner can be seen in Figure 8.

Figure 6. Broadcast for DWET

Figure 7. Broadcast for DZKB

Project name:	Thesis RadioPlanner
Customer:	
Data:	2020/04/01 11:33
Radio System Type:	Radio or TV Broadcasting
Propagation Model Type:	ITU-R P.1812-4
Percentage of time:	95%
Percentage of location:	95%
Margin:	0 dB
Mobile unit location:	Mobile unit with antenna below clutter height in urban or suburban environments
Clutter loss:	No
Area Study Type:	Field strength at remote
Rx antenna height:	10 m

	Field strength
	> 80 dBμV/m
	> 75 dBμV/m
	> 70 dBμV/m
	> 65 dBμV/m
	> 60 dBμV/m
	> 55 dBμV/m

Transmitter Parameters

№	Name	Latitude Longitude	Antenna azimuth	Antenna model	Antenna height	Antenna gain, dBi	Tx power, W	Loss, dB
1	DWGT	N14.654444° E121.045833°	0°	Kathrein 75010067	150 m	8.4	50000	1.6
2	DWET	N14.721945° E121.056945°	0°	Kathrein 75010067	200 m	8.4	60000	1.6
3	DZBB	N14.669617° E121.050061°	0°	Kathrein 75010067	236.8 m	8.4	100000	2.2
4	DZKB	N14.638333° E121.029722°	0°	Kathrein 75010067	228.6 m	8.4	50000	2.2
5	DZOE	N14.686283° E121.066728°	0°	Kathrein 75010067	236.8 m	8.4	100000	2.3
6	DZTV	N14.649444° E121.018889°	0°	Kathrein 75010067	200.5 m	8.4	50000	2.3
7	DWDB	N14.686283° E121.066728°	0°	Kathrein 75010067	236.8 m	8.4	30000	3.5
8	DWAO	N14.608056° E121.164722°	0°	Kathrein 75010067	202.5 m	8.4	60000	3.7

Figure 8. Data exported from RadioPlanner

The reinforcement learning algorithm

Pseudocode

//The Preparation Stage

- Gather secondary user data (i.e. current time, location)
- Import different channel data
- Initialise channel rewards
- Initialise q-table of size (number of channels, number of possible actions) // all elements are zeros
- Initialise algorithm parameters
  - number of episodes
  - steps per episode
  - learning rate
  - discount rate
  - exploration rate
  - max exploration rate
  - minimum exploration rate
  - exploration decay rate
- Initialise rewards per episodes array of size number of episodes // all elements are initially zero

//The Training Stage

- for episode in number of episodes
  - Set done variable to False
  - Set current reward to 0
  - for step in number of steps per episode
    - Generate random exploration rate threshold between 0 and 1
    - if exploration rate threshold is greater than exploration rate
      - Exploit, choose action with highest q-value from q-table
      - Get reward from action done
    - else
      - Explore, choose random action
      - Get reward from action done
    - Check if done is true depending on whether action causes end of episode
      - Update q-table
      - If done is true, break the loop and start new episode
    - Update exploration rate
    - Assign current reward of episode to current element in rewards table
    - End of for loop (steps)
- End of for loop (episodes)

//The Testing Stage

- Based on the final q-table, get the best action (max q-value)

A Q-table was created, which was composed of a 4x2 matrix of zeros. Row (4) represents the states, or the four channels, while column (2) represents the actions of choosing the availability. The algorithm parameters, which were composed of the number of episodes, steps per episode, learning rate, discount rate, exploration rate, maximum exploration rate, minimum exploration rate, and exploration decay rate, were all defined. For the actual training of the algorithm, a loop was used to divide each training into an episode.

The training began by choosing a random number between 0 and 1 for the exploration rate threshold. If the exploration rate threshold was greater than the predefined exploration rate, it would follow the exploitation theory and obtain the maximum q-value from the Q-table.

If the reward was equal to 100 or -100, then the done action would be equal to 1, which would be used later on. Otherwise, it would follow the exploration theory and obtain random elements from the Q-table action section. If the reward was equal to 100 or -100, done would be equated to 1 again. Afterward, the Q-table would be updated using the formula. The current episode reward would be updated with the reward from the action. If done was equal to 1, then it would break from the loop. The exploration rate would be updated using the formula, and the current episode reward would be updated to the list of rewards.

The first stage of the RL process was to set up the SU inputs. These inputs determined the reward table where the reward was dependent on how right a channel was for the SU. The reward table consisted of size (number of channels, number of possible actions). In this case, there were two possible actions of each channel, to accept it or not. Accepting and not accepting had their corresponding rewards. Thus, the most optimal channel would get the highest reward for accepting and the lowest reward for rejecting. Other reward values varied.

The next stage was to set up the environment, which in this case, ws the database. The database was generally non-linear for channels that were split into different blocks. In addition to the database, the Q-table was initiated with all zero values primarily. This table had the same size as the reward table. The Q-table was the basis for which actions were the best for any given state, based on the q-values.After setting up the environment, the algorithm parameters were initialised, which were used in the learning algorithm. The number of episodes determined how many learning sessions the agent had to do. The steps per episode were the maximum number of actions taken per episode. Learning rate and discount rates were variables used for updating a q-value in a Q-table.In RL, the agent either explored or exploited the environment. Certainly, the best was to have a balance between the two. This point was where the balancing parameters come in. Initially, the exploration rate was at 1, the maximum value. This value meant that the agent explored every learning process for every start because there was nothing to exploit yet. The

minimum exploration rate was simply the lowest possible exploration rate as the exploration rate decayed per episode. These values were flexible and could be changed, which might lead to better results.

Moving on to the learning process itself, a table of the rewards each episode finished was initialised. This table differed from the previous reward table, which had rewards for each action taken (not the agent's reward for a single episode).

The whole learning process was inside a big loop. The number of loops was the number of episodes that the agent took action. There was a done indicator signifying whether the agent committed a mistake or reached the goal of ending the episode in this loop. The current reward was also initialised. This position was where the current episode reward was stored first.

Inside the giant loop was another loop. This loop was based on the number of steps per episode. A step meanth a single action. Inside this loop was where the agent determined whether it explored or exploited. This action was based on the exploration rate and exploration rate threshold, whereby the threshold was randomized while the exploration rate decayed per step. In this step, the agent either explored, chose a random action, or exploited, choosing the best-known action so far. Both then returned an award, which determined how well the action was. Based on the current reward, current state, learning rate, current q-value, and discount rate, the q-value of the current action-state pair was then updated in the q-table. The current reward was also stored in the episode-reward table.

If the agent indeed committed a fatal mistake or somehow found the best channel, the episode would end. If not, then the agent could continue to take action until the maximum number of actions was made. Everything mentioned so far described the learning process. This process went on until the last episode.

After the learning process was finished, the test run involved the agent getting the highest q-value per action-state pair. In this case, the action in a corresponding state with the highest q-value was equivalent to choosing the best channel. In simpler terms, the action of taking channel x as the most optimal channel for the SU had the highest q-value, which meant that it was the best action.

Results and Discussion:

To further analyse the outcome of the Q-table, for every state (channel), there were two q-values. One q-value was for accepting the channel, the other for rejecting. If the q-value of accepting a channel was higher than the q-value of rejecting a channel, this siginfied that the SU could use that channel. However, it might not be the best depending on whether it was the highest q-value in all of the Q-table and vice versa. The Q-table determined the best action and at the same time, showed all other corrective actions for every channel (state).

Figure 9 shows predefined rewards for channels 11, 4, 5, 7, and 9. The reward table was divided into five rows and two columns. The rows represented the channels while the columns represented the states deciding to choose the channel. Channel 11 had the most available channels; therefore if the agent decided to choose it, it was rewarded with a value of 200. If the agent did not choose it, it was punished with a value of -200 because it did not choose an ideal state. Channels 4, 7, and 9 all had similar channel availabilities, which was why they had the same reward values. The agent should not choose these because they did not have a good number of available channels; therefore if the agent chose it, it was punished with a reward of -100. If the agent decided not to choose these channels, it was rewarded with a value of 100. Lastly, channel 5, which was located with [100 -100] reward values, had more available channels than channels 4, 7, and 9, but less than channel 11. If the agent decided to choose this channel, it was rewarded with a reward value of 100.

```
ch_rewards = 5×2
   200  -200
  -100   100
   100  -100
  -100   100
  -100   100
```

Figure 9. Predefined rewards for channels 11, 4, 5, 7, and 9

Figure 10 shows the Q-table that represents the actions of accepting and rejecting a channel, accepting on the left, and rejecting the right. A higher value in the accepting portion indicated that the user would use the said channel. Moreover, having a higher acceptance value signified that these channels should be selected because the agent was rewarded for it.

```
q_table = 5×2
10^4 ×
    2.0000    1.9600
    1.9700    1.9900
    1.9900    1.9700
    1.9700    1.9900
    1.9700    1.9900
```

Figure 10. Q-table after training

Figure 11 shows the output that provided the test action or the highest q-value in the Q-table. The channel with the highest Q-value would be provided to the user as it had the greatest availability of the other channels. The test_row and test_col were merely values that gave the index of the channel in the Q-table. In this case, channel 11 was located with an index row and column of [1,1]. Lastly, using the tic and toc MATLAB functions, the elapsed time for running the code was produced.

```
test_action = 2.0000e+04
test_row = 1
test_col = 1
Elapsed time is 0.043131 seconds.
```

Figure 11. Highest q-value, the index of the q-value, and the time taken to locate it

The code was created to output the most optimal channel depending on the user's time to access the database. The code obtained the time they logged in and would round that up to the next 30 minutes. This process would prevent the user from obtaining a time that would be wasted. The code would then look at the different channel databases to see which of them had the highest number of consecutive 1s. This checking would be considered the most optimal because it would provide the user with the longest amount of time to access the white space for personal use. The training was done to assure that the agent would obtain the greatest number of 1s. The results in the Q-table defined the best channel that the user would use. This channel would be provided using the max function in the first column as this was the column that would produce a choose action. Having a better value in the chosen column than in the deny column in the Q-table meant that the agent preferred to choose the channel rather than deny it. The Q-table also proved that the agent could learn to choose the best channel and if the agent had to accept or deny. In a more updated code, the researchers trained the agent, and it gave channel 11 as the best channel that could be provided.

Conclusion:

The use of TVWS in communication can address the problem of spectrum scarcity. One of the promising techniques in its implementation is with the use of TVWSDB. Its main purpose is to protect PUs from interference with SUs. There are several geolocation databases available for this purpose. However, it is unclear if those databases have the prediction feature that gives TVWSDB the capability of decreasing the number of inquiries from SUs. At present, there are very few studies on the implementation of RL in TVWS. This study sheds light on implementing AI, specifically RL, in TVWS. The benefit of this is that the agent can explore the environment through trial and error, and the information obtained from said exploration can be used to improve the overall output. Furthermore, this paper presents the advantage of using a machine learning approach in TVWSDB with an accurate and faster-searching capability for the available TVWS channels intended for SUs.

A user interface showing the contour protection maps of PUs and SUs would be a valuable feature to highlight the efficiency and effectiveness of TVWSDB in terms of interference detection.

Acknowledgment:

The authors acknowledge De La Salle University for its support in the publication of this paper.

Authors' declaration:

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Besides, the Figures and images, which are not ours, have been given the permission for re-publication attached with the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in De La Selle University.

References:

1. Prata A, Oliveira ASR, Carvalho NB. An Agile Digital Radio System for UHF White Spaces. IEEE Microwave Magazine. 2014;15,92–97.
2. Barrie M, Delaere S, Sukarevičienė G, Gesquiere J, Moerman I. Geolocation database beyond TV white spaces? Matching applications with database requirements. 2012 IEEE International Symposium on Dynamic Spectrum Access Networks. 2012;467–478.
3. Luo Y, Gao L, Huang J. Business modeling for TV white space networks. IEEE Communications Magazine, 2015;53,82–88.
4. Trinidad E, Materum L. Juxtaposition of Extant TV White Space Technologies for Long-Range Opportunistic Wireless Communications. International Journal of Emerging Trends in Engineering Research. 2019.
5. Pakzad AE, Pakzad AA, Materum L. Proposed Joint Propagation and Reinforcement Learning-based Television White Space Ledger. International Journal of Emerging Trends in Engineering Research. 2020;8(4).
6. Anabi KH, Nordin R, Abdullah NF. Database-assisted television white space technology:

Challenges, trends and future research directions. 2016;4,8162–8183.
7. Mueck M, Noguet D. TV white space standardization and regulation in Europe. 2011 2nd International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology (Wireless VITAE). 2011;1–5.
8. Feng X, Zhang J, Zhang Q. Database-assisted multi-ap network on tv white spaces: Architecture, spectrum allocation and ap discovery. 2011 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), 2011;265–276.
9. Sun C, Villardi GP, Lan Z, Alemseged YD, Tran HN, Harada H. Coexistence of secondary user networks under primary user constraints in TV white space. 2012 IEEE Wireless Communications and Networking Conference (WCNC). 2012;2146–2150.
10. Kokkinen H. TV White Space Spectrum Sharing Using Geolocation Databases. In TV White Space Communications and Networks. Elsevier. 2018;29-43.
11. Ojaniemi J, Poikonen J, Wichman R. Effect of geolocation database update algorithms to the use of TV white spaces. 2012 7th International ICST Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM). 2012;18–23.
12. Xue Z, & Wang L. Geolocation database based resource sharing among multiple device-to-device links in TV white space. 2015 International Conference on Wireless Communications & Signal Processing (WCSP). 2015;1–6.
13. Puspita RH, Shah SDA, Lee G, Roh B, Oh J, Kang S. Reinforcement Learning Based 5G Enabled Cognitive Radio Networks. 2019 International Conference on Information and Communication Technology Convergence (ICTC). 2019;555–558.
14. Dionisio R, Ribeiro J, Ribeiro J, Marques P, Rodriguez J. Cross-platform demonstrator combining spectrum sensing and a geo-location database. 2012 Future Network & Mobile Summit (FutureNetw), 2012;1–9.
15. Martin J, Dooley L, Wong K. A new cross-layer dynamic spectrum access architecture for TV white space cognitive radio applications. 2013.
16. Alhammadi A, Roslee M, Alias MY. Fuzzy logic based negotiation approach for spectrum handoff in cognitive radio network. 2016 IEEE 3rd International Symposium on Telecommunication Technologies (ISTT). 2016;120–124.
17. Wang C, Ma M, Zhao Z. Design of a novel dynamic trust model for spectrum management in WRANs of TV white space. Journal of Network and Computer Applications. 2017;100,1–10.
18. Sutton RS, Barto AG. Reinforcement learning: An introduction. MIT press. 2018.
19. Mathworks. (n.d.). Reinforcement Learning with MATLAB Understanding the Basics and Setting Up the Environment the Environment.
20. Government of Canada, DBS-01 — White Space Database Specifications. 2020. available at https://www.ic.gc.ca/eic/site/smt-gst.nsf/eng/sf10928.html.
21. Zurutuza, N. Cognitive Radio and TV White Space Communications: TV White Space Geolocation Database System [Master's Thesis]:Institutt for elektronikk og telekommunikasjon. 2011.

تعزيز قاعدة بيانات الفضاء الأبيض للتلفزيون القائم على التعلم

أرمي إي باكزاد رين ماتيوس مانويل جيريك سبنسر أوي
كزافييه فرانسيس أسونسيون جوشوا فينسينت ليجايو لورانس ماتيروم

1 جامعة دي لا سيل ، الفلبين.
2 جامعة مدينة طوكيو ، اليابان.

الخلاصة:
تشير المساحات البيضاء في التلفزيون (TVWSs) إلى الجزء غير المستخدم من الطيف تحت نطاقي الترددات العالية جدًا (VHF) والتردد الفائق TVWS .(UHF) هي ترددات تخضع لمستخدمين أساسيين مرخصين (PUs) لا يتم استخدامها ومتاحة للمستخدمين الثانويين .(SU)وهناك عدة طرق لتطبيق TVWS في الاتصالات ، من بينها استخدام قاعدة بيانات (TVWSDB) TVWS .وان الغرض الأساسي من TVWSDB هو حماية PU من التداخل مع وحدات التخزين. وهناك العديد من قواعد بيانات تحديد الموقع الجغرافي المتاحة لهذا الغرض. ومع ذلك ، ليس من الواضح ما إذا كانت قواعد البيانات هذه تتمتع بميزة التنبؤ التي تمنح TVWSDB القدرة على تقليل عدد الاستفسارات من وحدات النظام. مع وضع هذا في الاعتبار ، يقدم المؤلفون TVWSDB القائمة على التعلم المعزز. التعلم المعزز (RL) هو أسلوب للتعلم الآلي يركز على ما تم القيام به بناءً على تعيين المواقف للإجراءات للحصول على أعلى مكافأة. تم إجراء عملية التعلم من خلال تجربة الإجراءات للحصول على المكافأة بدلاً من إخبارك بما يجب القيام به. وقد تؤثر الإجراءات بشكل مباشر على المكافآت والمكافآت المستقبلية. استنادًا إلى النتائج ، قامت هذه الخوارزمية بالبحث بشكل فعال عن القناة الأكثر مثالية لوحدات SU في الاستعلام بأقل مدة بحث. ويقدم هذا البحث ميزة استخدام نهج التعلم الآلي في TVWSDB مع إمكانية بحث دقيقة وأسرع لقنوات TVWS المتاحة والمخصصة للنظم الخاصة.

الكلمات المفتاحية: الانتشار الراديوي ، إدارة الطيف الراديوي ، التعلم المعزز ، قاعدة بيانات الفضاء الأبيض للتلفزيون