electronics

Article

Decision Making for Communication Anti-Jamming Tasks with Knowledge-Graph-Based Q-Learning

Xijin Feng [1,2], Yingtao Niu [2,*], Qi Liu [1,3] and Quan Zhou [4]

[1] School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China; fx2201@163.com (X.F.)
[2] The Sixty-Third Research Institute, National University of Defense Technology, Nanjing 210007, China
[3] Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing 210044, China
[4] College of Communications Engineering, Army Engineering University of PLA, Nanjing 210007, China; zq_fnn@foxmail.com
* Correspondence: niuyingtao78@hotmail.com

Abstract: Due to the severe threats posed by smart jammers, anti-jamming decision making has become an essential technology for wireless communications. Most of the existing anti-jamming decision-making approaches have adopted Q-Learning to improve accuracy. However, the performances of these approaches drop dramatically in fast-varying jamming environments. Thus, an advanced Q-Learning approach utilizing domain knowledge graph as prior knowledge is proposed to select the optimal strategies with high flexibility and accuracy in different jamming environments. Specifically, by taking a knowledge graph that contains anti-jamming knowledge to initialize the Q-table, Q-Learning can avoid becoming stuck at local suboptimal solutions and obtain accurate strategies with fewer iterations. The iterations of the proposed approach are one third of those of other approaches based on Q-Learning and the average rewards of the proposed approach have improved by 2 percent. Numerical results demonstrate the optimality and excellent performance of the proposed approach over various existing benchmarks.

Keywords: wireless communications; anti-jamming; knowledge graph; Q-Learning; reinforcement learning

check for updates

Citation: Feng, X.; Niu, Y.; Liu, Q.; Zhou, Q. Decision Making for Communication Anti-Jamming Tasks with Knowledge-Graph-Based Q-Learning. *Electronics* **2024**, *13*, 4757. https://doi.org/10.3390/electronics13234757

Academic Editor: Dimitra I. Kaklamani

Received: 4 November 2024
Revised: 29 November 2024
Accepted: 29 November 2024
Published: 2 December 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Recently, wireless communication has become a vital technology for promoting social development and improving people's daily lives [1–4]. However, wireless communication systems are faced with complex, time-varying, and unsuspected electromagnetic environment due to the openness of wireless communication channels. Wireless communication systems are particularly vulnerable to external malicious jamming in many realistic application scenarios [5–7], which negatively impacts their reliability and effectiveness [8–10]. For example, in industrial applications, jamming can severely disrupt machine-to-machine communication. This disruption may halt production lines, cause downtime, and result in significant financial losses. Such interference highlights the critical need for reliable wireless communication to maintain operational efficiency and productivity in modern industrial systems.

To address these challenges, many works aim to reduce or eliminate the influence of interference on wireless communication systems by technical means, so as to ensure the reliability and effectiveness of communication, commonly referred to as communication anti-jamming. Traditional anti-jamming approaches focus on altering signal properties or exploiting specific channel characteristics to mitigate or avoid the effects of jamming, such as time-hopping and the use of ultrawideband signals; these are effective against fixed-jamming strategies [11,12]. However, their performances drop dramatically when treating

variable-jamming strategies, such as random pulse and intelligence barrage jamming. Furthermore, malicious jamming is becoming diversified, dynamic, and intelligent [13,14]. In this regard, wireless communication systems must be provided with further intelligent anti-jamming capabilities.

In recent years, machine learning and deep learning have been widely utilized in various fields [15–18]. With the great achievements of machine learning in the communication field, researchers have explored machine learning for efficient intelligent anti-jamming approaches [19–21]. As a kind of machine learning, reinforcement learning is capable of learning regular patterns of environmental changes without prior knowledge and is expected to achieve intelligent anti-jamming communication approaches. In dynamic jamming environments, wireless communication systems utilize appropriate actions which are decided by reinforcement learning algorithms to improve communication transmission efficiency and reliability [22–25].

Previous studies have utilized reinforcement learning technologies to optimize decision-making approaches, which improve anti-jamming decision-making accuracy with low algorithm complexity. However, most of the existing approaches ignore two challenging problems. On the one hand, communication systems cannot flexibly select optimal anti-jamming strategies according to different jamming environments, thus resulting in the system needing to relearn when the jamming environment changes. On the other hand, existing wireless communication knowledge is hardly adopted to improve the decision-making efficiency [26].

This paper explores the effects of utilizing anti-jamming prior-knowledge-based knowledge graphs to improve the convergence speed for reinforcement learning. A knowledge graph is comprised of entities, relationships, and semantic descriptions. Entities can be either concrete objects or abstract concepts, while relationships capture the connections between these entities. Semantic descriptions, on the other hand, are composed of types and properties that have specific meanings [27,28]. Domain-specific knowledge graphs are more complex and have a much richer knowledge structure that play an important role in real applications [29].

This paper utilizes a domain knowledge graph [30] which contains communication-based professional knowledge for improving the decision-making efficiency of Q-Learning. The proposed knowledge-graph-based Q-Learning (KGBQ) approach adequately utilizes the communication anti-jamming knowledge as prior knowledge. This approach initializes a Q-table by prior knowledge; thus, it does not require relearning when the jamming environment changes, enhancing its practicality. Firstly, perceived jamming features are matched in the domain knowledge graph. Then, the corresponding decision strategy in the knowledge graph is adopted to initialize the Q-value. Commonly used terms in the field covered by this paper are detailed in Table 1.The main contributions of this paper are as follows:

1. A novel anti-jamming decision-making approach based on reinforcement learning is proposed to improve the decision-making efficiency in unsuspected electromagnetic environment.
2. A domain knowledge graph is applied to Q-Learning for the first time. The proposed approach effectively uses prior knowledge for improving the performance of anti-jamming decision making.
3. Some experiments, based on more challenging jamming scenarios, are performed in this paper. The experimental results prove that the proposed approach can achieve an advanced performance.

The rest part of this paper proceeds as follows: Section 2 provides a literature review of relevant works; Section 3 lays out the system model and problem formulation; Section 4 is concerned with the algorithm used for this study; Section 5 analyses the results of the simulation; Section 6 gives a summary and future prospects.

Table 1. Commonly used terms and abbreviations in the paper.

Full Name	Abbreviation	Description
Reinforcement learning	RL	A machine learning method where an agent learns to take actions by interacting with the environment to maximize cumulative rewards.
Deep Q-Learning	DQL	A reinforcement learning method that combines deep neural networks with the Q-learning algorithm.
Markov Decision Process	MDP	A mathematical framework for modeling decision making, where outcomes depend solely on the current state and action, following the Markov property.
Acknowledgement (ACK) signal	ACK signal	A feedback signal sent by a receiver to the transmitter to confirm successful reception of a signal sent by the legitimate transmitter.
Knowledge-graph-based Q-Learning	KGBQ	An approach proposed in this paper

2. Related Works

This section is divided into two parts. The first part provides some applications of Q-Learning approaches for communication anti-jamming. The second part describes domain knowledge graphs and their applications.

2.1. Q-Learning-Based Communication Anti-Jamming

Reinforcement learning (RL) is a sub-branch of machine learning that is capable of learning regular patterns of environmental changes without prior knowledge [31]. Q-Learning is a specific algorithm in RL that learns the optimal strategy by estimating the value of state–action pairs, which is expected to achieve intelligent anti-jamming communication.

In [22], Yao et al. adopted the Markov game framework to propose a Q-Learning-based collaborative multi-agent anti-jamming algorithm (CMAA), which not only tackles the external malicious jamming, but also copes with the mutual interference among users. Elleuch et al. [23] proposed Cross-Check Q-Learning to help users learn their environment and the attacker's policy, effectively avoiding mutual interference. Zhou et al. [24] proposed a Q-Learning-based joint multi-agent anti-jamming algorithm (JMAA) to avoid internal mutual interference among sensor nodes. Zhou et al. [32] proposed a time–domain anti-pulse jamming (TDAA) algorithm that models the anti-jamming problem as a Markov decision process, making the agent continuously learn from the dynamic interaction with the jamming environment. Li et al. [33] adopted deep Q-Learning (DQL) to select communication frequency and power for achieving lower power and high throughput. Xu et al. [34] designed a two-channel Q-Learning structure to simulate the adversarial between jammer and cognitive radio network to improve the generalization of jamming attack defense. In [35], Chang et al. carried out a random jamming patterns switching strategy to improve the performance of DQL in a dynamic jamming environment. This approach obtains better performance and lower computational complexity in anti-jamming, and thus we compare the proposed KGBQ with this approach in the simulation experiment. However, these approaches have not utilized the existing wireless communication knowledge and are still poor in computation efficiency. This paper utilizes the anti-jamming prior knowledge to constructing the knowledge graph that aims to improve the convergence speed of Q-Learning. Table 2 also summarizes the above works.

Table 2. Comparison of Q-Learning-based communication anti-jamming approaches.

Reference No.	Year	Algorithm	Jamming Pattern	Domain Involved
[22]	2019	CMAA (Q-Learning)	sweep jamming	frequency
[23]	2021	Cross-Check Q-Learning	sweep jamming	frequency
[24]	2021	JMAA(Q-Learning)	blocking jamming	frequency
[32]	2020	TDAA(Q-Learning)	pulse jamming	time
[33]	2019	DQL	sweep jamming	frequency and power
[34]	2020	Two-channel Q-Learning	sweep jamming, random jamming, sensing-based jamming	frequency
[35]	2022	DQL	sweep jamming	frequency

2.2. Domain Knowledge Graph

A domain knowledge graph, also known as an industry or vertical knowledge graph, is built upon domain-specific data. A domain knowledge graph differs from a general knowledge graph, which includes common sense information. The information in a domain knowledge graph is primarily tailored to a specific field. They typically contain more complex knowledge and structures, making them crucial for real-world applications [36].

The first hypertension medication hyper-relational knowledge graph based on Chinese hyper-tension guidelines was proposed in [37] to provide medication decision support. The knowledge graph of COVID-19 infectors activity information has been constructed and used in [38]. Tang et al. [39] proposed an engineering-based approach that can improve upon traditional natural language processing to construct a petroleum exploration and development ontology domain knowledge graph. However, these approaches have not adopted the knowledge graph into communication anti-jamming. The proposed KGBQ constructs the communication professional knowledge as a domain knowledge graph to improve the performance of Q-Learning-based anti-jamming. Noticing that the optimization process of DQL still lacks explainability, results in the difficulty of mapping the knowledge graph into the deep features. Therefore, we adopt the knowledge graph to improve the traditional Q-Learning rather than deep Q-Learning

3. System Model and Problem Formulation

This proposed approach is based on Markov decision process (MDP) to model a single-agent reinforcement learning problem; this section details the system model and problem formulation.

3.1. System Model

A wireless communication environment contains a jammer, a transmitter, and a receiver. As shown in Figure 1, the legitimate transmitter transmits a signal; the corresponding receiver feeds back an acknowledgement (ACK) signal to the transmitter when the signal is successfully received.

This paper makes following assumptions for problem formulation:

1. An agent is located at the receiver side to explore jamming features. It utilizes the inserted knowledge graph to assist the receiver in generating anti-jamming decisions, and then feeds the decisions back to the transmitter.

2. The jammer interferes and destroys wireless communication transmissions by adjusting the jamming channel.

3. The transmission/jamming time is divided into equilong timeslots as the minimum time unit for continuous transmission, jamming, and channel switching. In addition, the jamming will not transform during a timeslot.

4. The jamming periods consist of continuous communication timeslots, and the jamming regulation in each jamming period is invariable. The available channel set in the wireless communication system is $\mathrm{M} = \{1, 2, \ldots, M\}$; thus, the jammer will select one or more channels to interfere with the transmission information during per timeslot.

5. Each timeslot contains transmission sub-timeslot, sensing the sub-timeslot and the learning sub-timeslot. The legitimate transmitter sends a packet in the transmission sub-timeslot, and receives an ACK message if the signal is successfully transmitted. The sensing sub-timeslot is utilized to sense the channel which is interfered with by the malicious jamming process. The learning sub-timeslot is utilized to learn Q-Learning algorithms and generate transmission decisions.

We use a knowledge graph to initialize the Q-table for improving the Q-Learning, where s, r, and a represent state, reward, and action, respectively.

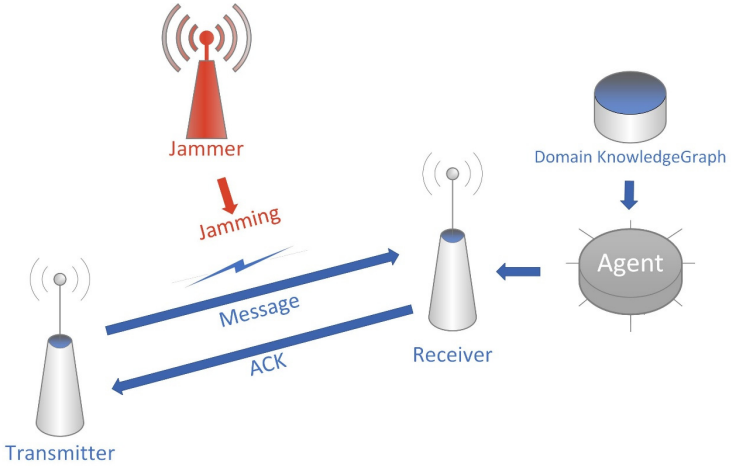


Figure 1. System model.

3.2. Problem Formulation

Anti-jamming problem in this paper is modeled as a Markov decision process (MDP). The MDP is defined by a four-tuple, $(\mathcal{S}, \mathcal{A}, P, r)$, where $\mathcal{S}$ is state space, $\mathcal{A}$ is action space, P is state transition probability, and r is the immediate reward of the system.

The state s_t and state space $\mathcal{S}$ are formulated as:

$$s_t \in S, S \triangleq \{1, 2, \cdots, M\} \tag{1}$$

where s_t is the jamming channel number $f_{j,t}$ at time t.

The action a_t depends on user's channel selection at time t; action space $\mathcal{A}$ can be formulated as:

$$a_t \in \mathcal{A}, \mathcal{A} \triangleq \{1, 2, \ldots, M\} \tag{2}$$

where a_t is the transmission channel number $f_{u,t}$ at time t.

The immediate reward r_t means the feedback of performing action a_t in the state space s_t, which depends on whether there is jamming signal in the selected transmission channel. In addition, the cost of channel switching is also considered in this paper:

$$r_t = \begin{cases} -1 - c_f * \varphi(f_{u,t}, f_{u,t-1}), & f_{u,t} = f_{j,t} \\ 1 - c_f * \varphi(f_{u,t}, f_{u,t-1}), & f_{u,t} \neq f_{j,t} \end{cases} \tag{3}$$

where $\varphi(f_{u,t}, f_{u,t-1})$ represents the channel-switching cost; c_f represents the channel-switching cost coefficient.

$$\varphi(f_{u,t}, f_{u,t-1}) = \begin{cases} 0, & f_{u,t} = f_{u,t-1} \\ 1, & other \end{cases} \tag{4}$$

When the user's transmission channel in current slot is same as that in previous slot, the cost of channel switching $\varphi(f_{u,t}, f_{u,t-1})$ is 0; otherwise, it is 1.

4. Knowledge-Graph-Based Q-Learning Algorithm

The proposed approach utilizes a domain knowledge graph to initialize a Q-table; the corresponding algorithm is as follows. The overall algorithm flow is detailed in Algorithm 1.

4.1. Knowledge Graph Construction

Domain knowledge graphs are hierarchically divided into a schema layer and a data layer. The schema layer serves as the logical foundation of the knowledge graph, typically employing an ontology model. Ontologies, as abstracted concepts, are used to describe the relationships, rules, and classifications of various entities. These ontologies are stored in the ontology model, encompassing elements such as jamming pattern categories and signal classifications. The data layer is formed through knowledge extraction and is mapped to the ontology model, representing concrete instances of the ontology. In the data layer, data are represented using triples in the form of "Entity (Attribute)—Relationship (Attribute)—Entity (Attribute).

The process of knowledge graph construction is as follows [26].

- Model the knowledge representation of jamming concepts, entities, attributes, relationships, and rules to form logical foundation of the jamming domain knowledge graph.
- Organize common signal types and jamming patterns as prior knowledge, and extract and process information on entities and attributes.
- Input the processed knowledge data into a graph database to construct the jamming domain knowledge graph.

The domain knowledge graph provides a structured representation of knowledge related to communication jamming, depicting entities and jamming strategies from a relational perspective. Based on prior knowledge of communication jamming, it performs knowledge extraction and processing for the domain, stores and visualizes this information in a graph database, and supports decision making in communication anti-jamming processes by enabling queries from the domain-specific knowledge graph.

4.2. Initialize Q-Table by Knowledge Graph

The agent in the communication system only senses a jamming signal without sending transmission instructions during the process of Q-value initialization. As shown in Figure 2, a Q-table combines the prior knowledge of domain knowledge graph and reinforcement learning modeling problem-based state–action space to initialize its value.

There are no direct corresponding anti-jamming strategies for unknown jamming patterns, and the factors determining corresponding anti-jamming strategies are different with the difference of unknown jamming patterns. How to use the anti-jamming knowledge graphs to make anti-jamming decisions for unknown jamming patterns is an important problem. The processing method of unknown patterns in this paper is as follows: firstly, process jamming parameters, and extract the time–frequency domain features of jamming patterns; then, search the time–frequency domain features and match them with their corresponding anti-jamming strategies [30]. The relationship from parameters to features can also be regarded as a triplet relationship, in which nodes are parameters and features, respectively, and edges are calculation rules.

Taking the periodic unknown dynamic jamming as an example, the features and parameters of the jamming signal are sensed during the initialization process. The jamming signal duration T_f of this unknown jamming pattern is relatively long and has periodicity; the temporary bandwidth is narrow. The jamming signal parameters include the initial frequency, f_L, and the terminate frequency, f_H.

The features of the jamming signal are matched with the existing features of the jamming patterns in the domain knowledge graph and then the decision-making strategy is decided to dynamic frequency adaptation. The decision-making parameters of this strategy are the frequency points of hopping and dwell time.

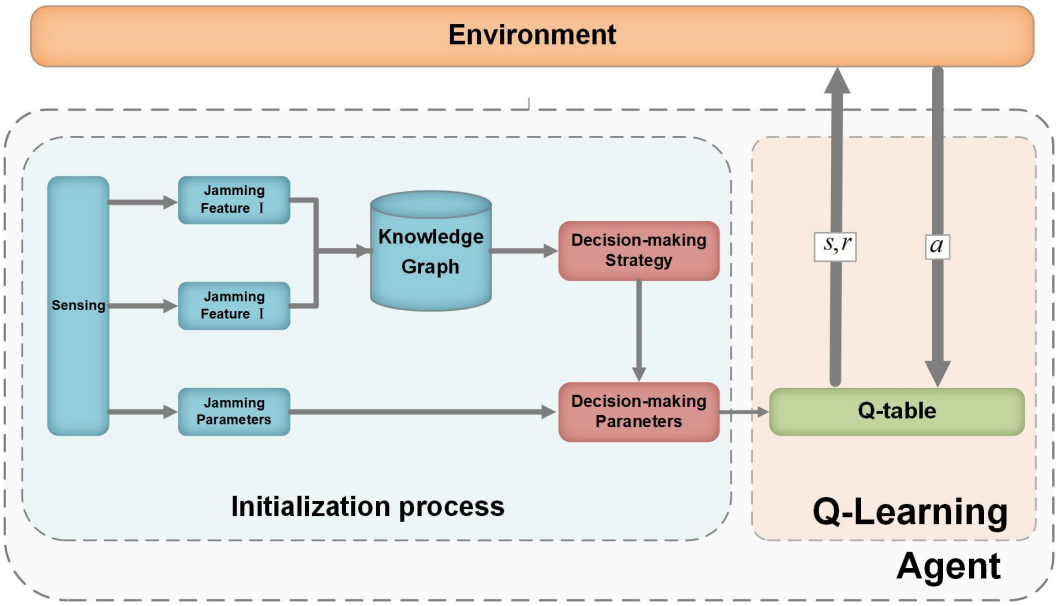


Figure 2. Schematic of knowledge-graph-based Q-Learning. The Q-table is obtained by combining decision-making strategy and parameters. The Q-table is initialized by knowledge graph to improve the Q-Learning performance, where s, r, and a represent state, reward, and action, respectively.

According to the initial and terminate frequency, f_L, f_H, and the jamming period, found by sensing the jamming signal, the decision-making parameters are calculated by the dynamic frequency adaptation decision-making strategy in the domain knowledge graph. The decision-making parameters are, respectively, frequency point of hopping $f_1 = \frac{3}{4}f_L + \frac{1}{4}f_H$, $f_2 = \frac{1}{4}f_L + \frac{3}{4}f_H$ and dwell time $T/2$.

The following values (i.e., 2, -1 or 0) based on decision-making parameters and jamming features are calculated by dynamic frequency adaptation strategy, and then utilize it as prior knowledge to initialize the Q-table. The initialized Q-table is defined as follows:

$$\mathcal{Q}(s_t, a_t) = \begin{cases} 2, & (a_t = f_1 \;\&\; t \le T/2) \text{ or } (a_t = f_2 \;\&\; t > T/2) \\ -1, & a_t = s_t \\ 0, & other \end{cases} \tag{5}$$

Specifically, the Q-value is assigned by high reward value of 2 when the action (transmission channel) is consistent with the hop frequency point. When the action is consistent with the state (jamming channel), the Q-value is assigned by the low reward value of 1. In other cases, the Q-value is 0.

4.3. Q-Learning

Due to the high optimization efficiency of TDAA [32], we adopt it to optimize the Q-values during the following transmission process.

The agent executes an action, a_t, based on the environment state, s_t, at time t by the current policy; then, it will obtain the corresponding reward, r_t, and update the state as s_{t+1}. This way, the iteration will be completed. According to the previous observation values and current reward, r_t, the Q-values are optimized as follows:

$$Q_t(s_t, a_t) = (1 - \alpha)Q_{t-1}(s_t, a_t) + ar_t \tag{6}$$

The agent executes the action of the last timeslot decision during the transmission sub-timeslot, and then senses the jamming channel during the sensing sub-timeslot. Finally, the agent utilizes the optimized Q-values to obtain the next timeslot action during the study

sub-timeslot and transmits the corresponding action to transmitter. When the Q-values do not fluctuate significantly, the Q-table can be regarded as having reached its optimum state.

4.4. Exploration and Exploitation

A fixed-ε value-greedy model is generally adapted by traditional Q-Learning algorithms for balancing exploration and exploitation. Fixed-ε greedy models have a probability $\varepsilon \in (0, 1)$ of randomly selecting an action in action space for action exploration. There is also a probability of $1 - \varepsilon$ that the model selects the best action at a given moment according to the Q-value of different actions in the current Q-table for exploiting existing information. In this way, this model learns the action selection of action space by the parameter ε.

In the communication anti-jamming scenario, the transmitter needs to utilize the anti-jamming decision making prior knowledge to establish stable communication behavior. However, the approach adopts the fixed value ε that balances the exploration, and exploitation is not conducive to the stable convergence of the learning process. This is because this approach has no consideration for utilizing prior knowledge. Thus, this paper realizes the smooth transition of exploration and exploitation based on the Softmax algorithm. The communication system utilizes complete anti-jamming decision-making knowledge to learn an intelligent communication anti-jamming approach. Thus, the proposed algorithm has a superior performance in stable communication. The communication selection strategy is updated as follows:

$$\pi(a_t|s_t) = \frac{e^{Q(s_t,a_t)/\xi}}{e^{Q(s_t,a_t)/\xi} + \sum\limits_{a^-} e^{Q(s_t,a^-)/\xi}} \tag{7}$$

where a_t represents the current communication channel and a^- represents other communication channels. The parameter ξ represents temperature that has negative linear correlation with the probability of selecting action with the larger Q-value. When the temperature ξ tends to zero, the agent will tend to select actions completely based on the Q-table. On the contrary, the agent will randomly select the action when the temperature ξ tends to infinity.

$$\xi = \begin{cases} \xi_0 e^{-vt} & \xi \geq \xi_{\text{final}} \\ \xi_{\text{final}} & \xi < \xi_{\text{final}} \end{cases} \tag{8}$$

The initial temperature ξ_0 has a positive correlation with the exploration ability in the initial stage of the algorithm. The parameter $v > 0$ can make the temperature ξ gradually approach zero with the algorithm iteration; thus, the proposed algorithm will gradually transition from the exploration stage to the exploitation stage. The values of v and ξ_{final}, respectively, determine the transition time and the end condition of the exploration phase.

Algorithm 1. Knowledge-Graph-based Q-Learning

1: **Initialization:**
   Sense the features and parameters of the jamming signal;
   Match the features of jamming patterns in domain knowledge graph;
   Obtain the decision-making strategy;
   Obtain the decision-making parameters;
   Initialize the Q-table.
2: **for** $t = 1, 2, \cdots, T$ **do**
3: The agent executes action a_t according to the environment state s_t;
4: Sense whether there is jamming at time t;
5: Obtain the corresponding reward r_t;
6: Update the Q-value (Formula (6));
7: The transmitter receives a new strategy by ACK;
8: **end for**

5. Simulation Experiment

The experimental results of the proposed approach with different parameters are reported, and then the simulation results are compared with other approaches by the determined parameter.

5.1. Parameter Determination

In this paper, the simulation experiment is carried out through MATLAB to evaluate the performance of the proposed KGBQ. The data of knowledge graph are organized by the Neo4j platform. The frequency band for data transmission and jamming is from 300 MHz to 380 MHz, which contains five non-overlapping 16 MHz channels. The detailed parameters of the simulation experiment are presented in Table 3.

Table 3. Parameter determination.

Quantity	Symbol	Value
Communication timeslot length	T_s	0.3 ms
Transmission sub-slot length	T_{tx}	0.02 ms
Sensing sub-slot length	T_{wbss}	0.004 ms
Learning sub-slot length	$T_{learning}$	0.006 ms
Number of communication timeslots	T_a	1200
Number of channels	M	16
Learning rate	α	0.24
Channel-switching cost coefficient	c_f	0.1

This paper carries out experiments against periodic unknown jamming; the jamming signal parameters are determined as shown in Table 4.

Table 4. Jamming signal parameter determination.

Quantity	Symbol	Value
Duration	T_f	0.36 s
Temporary bandwidth	B	1 MHz
Initial frequency	f_L	300 MHz
Terminate frequency	f_H	380 MHz
Signal period	T	4.8 ms

To evaluate the performance of the proposed approach, this section uses normalized throughput, average reward value, and channel-switching cost as metrics to compare the proposed KGBQ method with random frequency hopping (Rand) [12], Q-Learning [22] and DRL [23]. The performances of other algorithms (i.e., Rand frequency-hopping, Q-Learning, deep Q-Learning (DQL) [35]) and the proposed KGBQ are compared by adopting normalized throughput, average reward during iteration, and channel-switching cost as metrics. The Rand approach is widely used in traditional anti-jamming communication approaches due to its simplicity and low cost. It improves communication security and anti-interference capabilities by rapidly and randomly hopping the carrier frequency among several preset values. Normalized throughput is the amount of message that is successfully transferred per unit time and measures communication efficiency. Average reward is a common metric in reinforcement learning models to measure the effectiveness of learning [40]. The higher the reward value and the faster the reward value improvement rate, the better the performance. The channel-switching cost represents the loss of system performance caused by switching different communication channels, which can be calculated by Equation (4).

5.2. Simulation Results and Analysis

According to the parameters and features of unknown dynamic jamming, the detection anti-jamming strategy of the proposed algorithm is dynamic-frequency-adaptative, and its corresponding parameters are determined as shown in Table 5.

Table 5. Decision-making parameters.

Quantity	Symbol	Value
Frequency point of hopping	f_1, f_2	320 MHz 360 MHz
Dwell time	t	$T/2$

Figure 3 shows the performance of the convergence states of KGBQ, Q-Learning, and DRL methods under two different jamming approaches. The red areas represent jamming signals, while the other color areas represent user signals. When the user signal overlaps with the jamming signals, it means that the current strategy has not avoided communication jamming. As shown in Figure 3, compared to Q-Learning and DRL, the proposed KGBQ performs well in avoiding communication jamming. Specifically, the degree of overlap between the user signal region and the red interference region using the KGBQ method is significantly reduced, indicating that it can better select time–frequency resources and effectively avoid jamming.

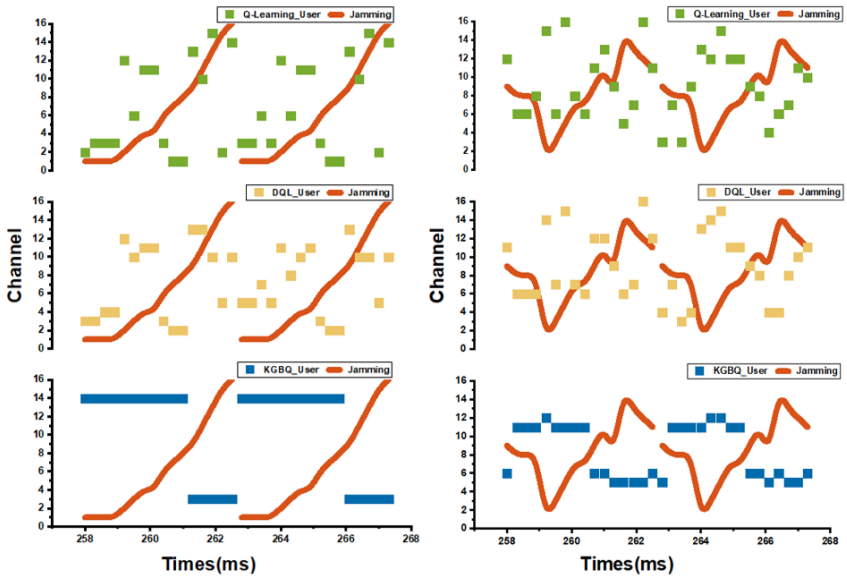


Figure 3. Jamming user signal time–frequency information.

Figure 4 shows the normalized throughput comparison of different approaches. It can be observed that the convergence efficiency of the proposed KGBQ is faster than those of the other three. Q-Learning and DQL need more than 500 iterations to converge, and the proposed KGBQ only needs about 100 iterations to converge. The major reason is that the proposed KGBQ utilizes the prior knowledge of the knowledge graph to initialize the Q-table, which makes convergence easier. Because the random frequency hopping approach has no sensing and learning ability, the normalized throughput is always around 0.9. The average reward value is calculated by Equation (3). A reward function guides an agent to complete an anti-jamming assignment efficiently. Figure 5 shows 1200 iteration datapoints, with every set of 10 iterations averaged to clearly illustrate the general trend. As shown in Table 6, the average reward value of the proposed solution increases significantly by about 10%. That is because prior knowledge can improve the action decision making of the agent; thus, the reward values are always high.

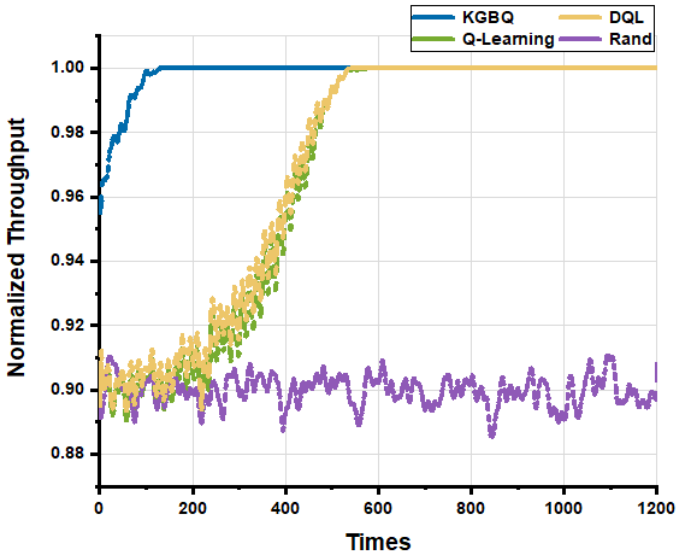


Figure 4. Normalized throughput.

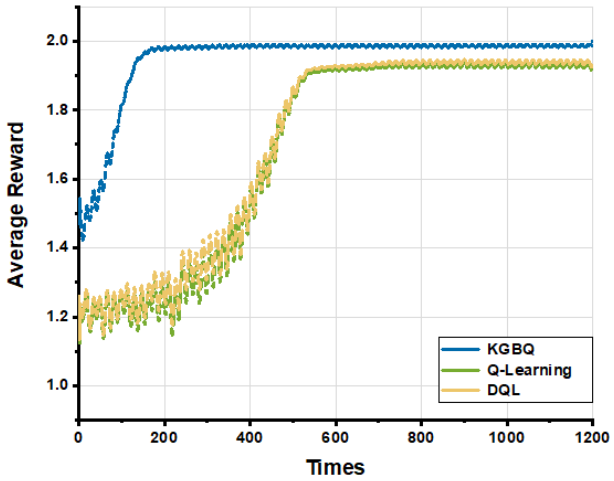


Figure 5. Comparison of average rewards.

Table 6. Comparison of rewards.

Approach	Iterations	Total Reward	Average Reward
KGBQ	1200	2340.339	1.948
Q-Learning	1200	1944.489	1.619
DQL	1200	2045.9005	1.703

Figure 6 also reports the comparison of channel-switching costs with the other three approaches. The reconstruction of the communication link increases the computational and time cost of the communication systems; thus, a user selects the different communication channel in the next timeslot (i.e., channel switching), which will cause a drop in system performance. The values of the channel-switching cost are calculated by Equation (4). The proposed KGBQ utilizes the knowledge graph containing much domain knowledge to initialize the Q-table; thus, the channel-switching cost of KGBQ is much lower than that of the others. As shown in Table 7, the average and total channel-switching costs are reported to evaluate the proposed approach and compare it with the three other solutions. This further demonstrates that the proposed algorithm has improved the performance in communication anti-jamming decision making. This is because the Q-table is initiated by

the proposed algorithm that contains prior knowledge to help users for a relatively fixed frequency hopping point.

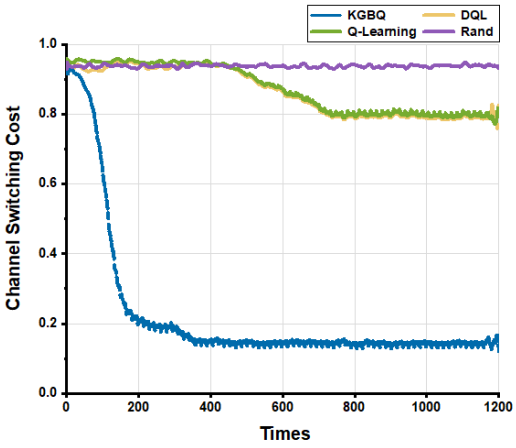


Figure 6. Comparison of channel-switching cost.

Table 7. Channel-switching cost.

Approach	Iterations	Total Channel-Switching Cost	Average Channel-Switching Cost
KGBQ	1200	268.805	0.223
Q-Learning	1200	1048.264	0.872
DQL	1200	1039.295	0.865
Rand	1200	1126.085	0.937

Figure 7 shows the superiorities of the proposed KGBQ method in various performances of communication anti-jamming decision making. It can be observed that the normalized throughput of KGBQ is infinitely close to 1. The average reward value of KGBQ during iteration is much higher than that of the other approaches. Furthermore, KGBQ achieves optimal performance in terms of the channel-switching cost.



Figure 7. Comparison value of different approaches.

6. Conclusions

In this paper, the KGBQ has been proposed to sufficiently utilize the prior knowledge to construct the domain knowledge graph for improving the performance on communica-

tion anti-jamming. Feature matching is employed to link the sensed jamming parameters with those in a knowledge graph, obtaining the corresponding anti-jamming decision-making strategy as domain knowledge. Then, this domain knowledge is integrated into Q-Learning, with the Q-table initialized using prior knowledge to boost learning performance. The experiment results demonstrate that KGBQ significantly improves anti-jamming decisions, increasing normalized throughput and decision-making efficiency.

The principal theoretical implication of this study is that the utilization of a domain knowledge graph can reduce algorithm complexity. The major limitation of the proposed method is that it is difficult to complete the optimal domain knowledge graph due to the constantly updated jamming patterns. Promising directions for exploration include, on one hand, leveraging existing domain knowledge graphs to reduce the state–action space in reinforcement learning, and on the other hand, integrating domain knowledge graphs with other approaches, such as reconfigurable intelligent surface (RIS) technology [41,42].

Author Contributions: Conceptualization, X.F. and Y.N.; methodology, X.F.; software, X.F.; validation, Q.L., Q.Z. and Y.N.; formal analysis, Y.N.; investigation, Q.Z.; resources, Q.L.; data curation, X.F.; writing—original draft preparation, X.F.; writing—review and editing, Y.N.; visualization, Q.Z.; supervision, Q.L.; project administration, Q.L.; funding acquisition, Y.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China grant number 62371461, National Key Research and Development Program of China grant number 2021 YFE0116900, Major Program of the National Social Science Fund of China grant number 17ZDA092, basic Research Programs (Natural Science Foundation) of Jiangsu Province grant number BK20191398; And The APC was funded by National Natural Science Foundation of China.

Data Availability Statement: Dataset available on request from the authors: the raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Qian, Y. Security and Wireless Communication Networks. *IEEE Wirel. Commun.* **2020**, *27*, 4–5. [CrossRef]
2. Ercan, A.; Mowla, M.N. Wireless communication protocols in smart agriculture: A review on applications, challenges and future trends. *Ad Hoc Netw.* **2022**, *136*, 102982.
3. Shaukat, K.; Alam, T.M.; Hameed, I.A.; Khan, W.A.; Abbas, N.; Luo, S. A review on security challenges in internet of things (IoT). In Proceedings of the 2021 26th International Conference on Automation and Computing (ICAC), Portsmouth, UK, 2–4 September 2021; pp. 1–6.
4. Javed, I.; Tang, X.; Shaukat, K.; Sarwar, M.U.; Alam, T.M.; Hameed, I.A.; Saleem, M.A. V2X-based mobile localization in 3D wireless sensor network. *Secur. Commun. Netw.* **2021**, *2021*, 6677896. [CrossRef]
5. Jia, L.; Qi, N.; Su, Z.; Chu, F.; Fang, S.; Wong, K.K.; Chae, C.B. Game Theory and Reinforcement Learning for Anti-jamming Defense in Wireless Communications: Current Research, Challenges, and Solutions. *IEEE Commun. Surv. Tutor.* 2024; *early access*.
6. Shaukat, K.; Alam, T.M.; Luo, S.; Shabbir, S.; Hameed, I.A.; Li, J.; Abbas, S.K.; Javed, U. A review of time-series anomaly detection techniques: A step to future perspectives. In *Advances in Information and Communication: Proceedings of the 2021 Future of Information and Communication Conference (FICC), Vancouver, BC, Canada, 29–30 April 2021*; Springer: Cham, Switzerland, 2021; pp. 865–877.
7. Shaukat, K.; Iqbal, F.; Hameed, I.A.; Hassan, M.U.; Luo, S.; Hassan, R.; Younas, A.; Ali, S.; Adeem, G.; Rubab, A.; et al. MAC protocols 802.11: A comparative study of throughput analysis and improved LEACH. In Proceedings of the 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Phuket, Thailand, 24–27 June 2020; pp. 421–426.
8. Pirayesh, H.; Zeng, H. Jamming attacks and anti-jamming strategies in wireless networks: A comprehensive survey. *IEEE Commun. Surv. Tutor.* **2022**, *24*, 767–809. [CrossRef]
9. Hassan, M.U.; Shahzaib, M.; Shaukat, K.; Hussain, S.N.; Mubasshir, M.; Karim, S.; Shabir, M.A. DEAR-2: An energy-aware routing protocol with guaranteed delivery in wireless ad-hoc networks. In *Recent Trends and Advances in Q5 Wireless and IoT-enabled Networks, EAI/Springer Innovations in Communication and Computing*; Springer: Cham, Switzerland, 2019; pp. 215–224.
10. Dar, K.S.; Hassan, M.U. Cloud Security through Encryption Techniques. *Transylv. Rev.* **2017**, *XXV*, 4037–4042.
11. Grover, K.; Lim, A.; Yang, Q. Jamming and anti-jamming techniques in wireless networks: A survey. *Int. J. Ad Hoc Ubiquitous Comput.* **2014**, *17*, 197–215. [CrossRef]
12. Chen, K.W.; Chao, C.M.; Lin, C.Y.; Yeh, C.C. Anti-jamming channel hopping protocol design based on channel occupancy probability for Cognitive Radio Networks. *Comput. Netw.* **2022**, *214*, 109125. [CrossRef]

13. Zhang, Y.; Jia, L. A fuzzy learning anti-jamming approach with incomplete information. *IEEE Commun. Lett.* **2024**, *28*, 1514–1518. [CrossRef]
14. Jia, L.; Qi, N.; Chu, F.; Fang, S.; Wang, X.; Ma, S.; Feng, S. Game-theoretic learning anti-jamming approaches in wireless networks. *IEEE Commun. Mag.* **2022**, *60*, 60–66. [CrossRef]
15. Zhou, Z.; Li, Y.; Li, J.; Yu, K.; Kou, G.; Wang, M.; Gupta, B.B. Gan-siamese network for cross-domain vehicle re-identification in intelligent transport systems. *IEEE Trans. Netw. Sci. Eng.* **2022**, *10*, 2779–2790. [CrossRef]
16. Zhou, Z.; Li, Y.; Zhang, Y.; Yin, Z.; Qi, L.; Ma, R. Residual visualization-guided explainable copy-relationship learning for image copy detection in social networks. *Knowl.-Based Syst.* **2021**, *228*, 107287. [CrossRef]
17. Shaukat, K.; Luo, S.; Varadharajan, V.; Hameed, I.A.; Xu, M. A survey on machine learning techniques for cyber security in the last decade. *IEEE Access* **2020**, *8*, 222310–222354. [CrossRef]
18. Shaukat, K.; Luo, S.; Varadharajan, V. A novel method for improving the robustness of deep learning-based malware detectors against adversarial attacks. *Eng. Appl. Artif. Intell.* **2022**, *116*, 105461. [CrossRef]
19. Hussain, F.; Hussain, R.; Hassan, S.A.; Hossain, E. Machine learning in IoT security: Current solutions and future challenges. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 1686–1721. [CrossRef]
20. Shi, Y.; Lu, X.; Niu, Y.; Li, Y. Efficient jamming identification in wireless communication: Using small sample data driven naive bayes classifier. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 1375–1379. [CrossRef]
21. Frikha, M.S.; Gammar, S.M.; Lahmadi, A.; Andrey, L. Reinforcement and deep reinforcement learning for wireless Internet of Things: A survey. *Comput. Commun.* **2021**, *178*, 98–113. [CrossRef]
22. Yao, F.; Jia, L. A Collaborative Multi-Agent Reinforcement Learning Anti-Jamming Algorithm in Wireless Networks. *IEEE Wirel. Commun. Lett.* **2019**, *8*, 1024–1027. [CrossRef]
23. Elleuch, I.; Pourranjbar, A.; Kaddoum, G. A novel distributed multi-agent reinforcement learning algorithm against jamming attacks. *IEEE Commun. Lett.* **2021**, *25*, 3204–3208. [CrossRef]
24. Zhou, Q.; Li, Y.; Niu, Y. Intelligent Anti-Jamming Communication for Wireless Sensor Net-works: A Multi-Agent Reinforcement Learning Approach. *IEEE Open J. Commun. Soc.* **2021**, *2*, 775–784. [CrossRef]
25. Zhang, X.; Wang, H.; Ruan, L.; Xu, Y.; Feng, Z. Joint Channel, Power and Bandwidth Optimization for Anti-jamming Communications: A Multi-agent Q-learning Approach. In Proceedings of the IEEE 13th International Conference on Wireless Communications and Signal Processing (WCSP), Changsha, China, 20–22 October 2021.
26. Niu, Y.; Feng, X.; Kou, S.; Xiang, P. A Novel Anti-Jamming Decision-Making Algorithm Based on Knowledge Graph Technology. *Appl. Sci.* **2022**, *12*, 4960. [CrossRef]
27. Ji, S.; Pan, S.; Cambria, E.; Marttinen, P.; Yu, P.S. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 494–514. [CrossRef] [PubMed]
28. Hogan, A.; Blomqvist, E.; Cochez, M.; Damato, C.; Melo, G.D.; Gutierrez, C.; Kirrane, S.; Gayo, J.E.L.; Navigli, R.; Neumaier, S.; et al. Knowledge graphs. *ACM Comput. Surv.* **2021**, *54*, 71:1–71:37.
29. Lin, J.; Zhao, Y.; Huang, W.; Liu, C.; Pu, H. Domain knowledge graph-based research progress of knowledge representation. *Neural Comput. Appl.* **2021**, *33*, 681–690. [CrossRef]
30. Kou, S.; Niu, Y.; Zhang, H. Construction of knowledge graph based on conventional jamming patterns. In Proceedings of the International Conference on Wireless and Satellite Systems (EAI ICWSS), Harbin, China, 31 July–2 August 2021; pp. 86–97.
31. Shaukat, K.; Luo, S.; Varadharajan, V.; Hameed, I.A.; Chen, S.; Liu, D.; Li, J. Performance comparison and current challenges of using machine learning techniques in cybersecurity. *Energies* **2020**, *13*, 2509. [CrossRef]
32. Zhou, Q.; Li, Y.; Niu, Y. A countermeasure against random pulse jamming in time domain based on reinforcement learning. *IEEE Access* **2020**, *8*, 97164–97174. [CrossRef]
33. Li, Y.; Xu, Y.; Wang, X.; Li, W.; Bai, W. Power and Frequency Selection optimization in Anti-Jamming Communication: A Deep Reinforcement Learning Approach. In Proceedings of the 2019 IEEE 5th International Conference on Computer and Communications (ICCC), Chengdu, China, 6–9 December 2019; pp. 815–820.
34. Xu, J.; Lou, H.; Zhang, W.; Sang, G. An Intelligent Anti-Jamming Scheme for Cognitive Radio Based on Deep Reinforcement Learning. *IEEE Access* **2020**, *8*, 202563–202572. [CrossRef]
35. Chang, X.; Li, Y.; Zhao, Y.; Du, Y.; Liu, D. An improved anti-jamming method based on deep reinforcement learning and feature engineering. *IEEE Access* **2022**, *10*, 69992–70000. [CrossRef]
36. Javed, U.; Shaukat, K.; Hameed, I.A.; Iqbal, F.; Alam, T.M.; Luo, S. A review of content-based and context-based recommendation systems. *Int. J. Emerg. Technol. Learn.* **2021**, *16*, 274–306. [CrossRef]
37. Zhou, G.; Haihong, E.; Kuang, Z.; Tan, L.; Xie, X.; Li, J.; Luo, H. Clinical decision support system for hypertension medication based on knowledge graph. *Comput. Methods Programs Biomed.* **2022**, *227*, 107220. [CrossRef]
38. Chen, L.; Liu, D.; Yang, J.; Jiang, M.; Liu, S.; Wang, Y. Construction and application of COVID-19 infectors activity information knowledge graph. *Comput. Biol. Med.* **2022**, *148*, 105908. [CrossRef]
39. Tang, X.; Feng, Z.; Xiao, Y.; Wang, M.; Ye, T.; Zhou, Y.; Meng, J.; Zhang, B.; Zhang, D. Construction and application of an ontology-based domain-specific knowledge graph for petroleum exploration and development. *Geosci. Front.* **2022**, *14*, 101426. [CrossRef]

40. Shaukat, K.; Luo, S.; Chen, S.; Liu, D. Cyber threat detection using machine learning techniques: A performance evaluation perspective. In Proceedings of the 2020 International Conference on Cyber Warfare and Security (ICCWS), Norfolk, VA, USA, 12–13 March 2020; pp. 1–6.
41. Zhi, K.; Pan, C.; Ren, H.; Chai, K.K.; Elkashlan, M. Active RIS versus passive RIS: Which is superior with the same power budget. *IEEE Commun. Lett.* **2022**, *26*, 1150–1154. [CrossRef]
42. Zhi, K.; Pan, C.; Ren, H.; Wang, K. Power scaling law analysis and phase shift optimization of RIS-aided massive MIMO systems with statistical CSI. *IEEE Trans. Commun.* **2022**, *70*, 3558–3574. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.