doi:10.11918/j.issn.0367-6234.2015.06.022

井下认知无线电 EWA 信道选择算法

孙 永,钱建生

(中国矿业大学 信息与电气工程学院,221008 江苏 徐州)

摘 要:为解决特殊的井下恶劣无线环境造成通信极易中断的问题,提出基于经验加权吸引策略的信道选择算法.深入分析报酬反馈结构,建立更加细致和合理的多级的报酬反馈体制.采用简单多次重复概率实验的方法验证了基于 EWA 学习的信道选择算法的有效性.通过与传统单一固定信道传输和 Q 学习算法的深入对比研究表明,基于 EWA 学习的信道选择算法根据无线信道特性的在线学习经验对优选信道快速捕获的准确性和稳定性方面具有更好的性能.

关键词:煤矿安全;无线通信系统;认知无线电;强化学习;无线信道;EWA

中图分类号:TN929.5 **文献标志码**:A **文章编号**:0367-6234(2015)06-0119-05

Channel EWA selection strategy in coal mine cognitive radio

SUN Yong, QIAN Jiansheng

(School of Information and Electrical Engineering, China University of Mining and Technology, 221008 Xuzhou, Jiangsu, China)

Abstract: In oder to solve extremely unstable link quality and communication failure caused by strong electromagnetic noise and interference in special coal mine environment, this paper proposes an Experience-Weighted Attraction(EWA) selection algorithm. Based on the system structure analysis in depth, it establishes more detailed and reasonable multi-layer reward feedback system. Research is conducted by the method of simple repeated probability experiments through software simulations. By comparing with traditional fixed single channel transmission and Q learning, channel selection strategy with EWA learning has more advantageous performance on sensitivity and stability.

Keywords: coal mine safety; wireless telcommunication systems; cognitive radio (CR); reinforcement learning(RL); communication channels; EWA(experience-weighted attraction)

煤矿井下目前存在多种无线通信系统如矿用小灵通通信、漏泄通信、中频感应通信、蜂窝通信、动力线载波通信以及超低频透地通信等[1].针对煤矿井下无线的研究和应用详见参考文献[2-7].然而不同井下无线通信系统的混合和有限无线资源[8-12]的激烈争夺,将导致重要通信系统(高优先级)如煤矿井下无线应急保障系统的无法通信和工作,严重危害煤矿井下的安全生产.同时在采煤工作面,随着采煤机的推进割煤的影响,无线环境的空间几何形状在不断变化,其无线信道特性也在不断变化,传统的单一信道链路质量极为不稳定易造成通信中断.具有智能感知无线通信环境并实时选择最优通信信道的认知无线电在煤矿井下的应用能最大程度地保证重要系统如煤矿井下无线应急保障系统的正常通信.

被认为是软件无线电的扩展和延伸的认知无线电的概念首先由文献[13]提出.作为智能无线电的革命性的发展,认知无线电在软件无线电的基础上加入了包含知识库,推理引擎和学习引擎在内的认知引擎,使得软件无线电能够学习和适应周围的无线环境[14].作为实现认知无线电的核心部件和关键功能,学习引擎负责通过积累新的环境经验,更新和扩展知识库,这也是区别于传统的预编程软件无线电的标志性特征.目前已有多种学习算法应用于认知无线电的学习引擎中,

收稿日期:2014-05-01.
基金项目:国家高技术研究发展计划(863 计划)(2012AA062103);
江苏省产学研创新基金前瞻性联合研究项目(BY2012081).
作者简介:孙 永(1985—),男,博士研究生;
钱建生(1964—),男,教授,博士生导师.
通信作者:钱建生,qianjsh@cumt.edu.cn.

包括强化学习算法在内的贝叶斯统计学习算法，基因遗传算法，神经网络算法[15-16]. 其中，以 Q 学习[17-22]为代表的强化学习研究最为众多和深入. 然而上述算法多对收敛点(如 Q 学习的纳什均衡点)或次优点的存在性进行较多的研究，而对现场应用中的收敛速率等实时性的实际问题缺乏更加深入的探讨. 基于 EWA 模型的信道选择算法能够使系统在线学习井下无线电环境通信信道特性，快速推测并选择和更改当前最优通信信道，实时动态保证通信链路质量，降低系统通信中断概率.

1 EWA 理论

EWA 模型源于多人普通博弈理论. 设共有 n 人参与博弈，每个参与者标记为 $i(i=1,2,\cdots,n-1,n)$. 选手 i 的策略空间为 $\boldsymbol{S}_i$，包含有 m_i 个离散的选项，即 $\boldsymbol{S}_i=\{s_i^1,s_i^2,\cdots,s_i^{m_i-1},s_i^{m_i}\}$. 参与者 i 的一个策略 s_i 是 $\boldsymbol{S}_i$ 的一个元素，即 $s_i\in\boldsymbol{S}_i$. 整个博弈的策略空间 $\boldsymbol{S}$ 是 n 个人策略空间的笛卡尔乘积，即 $\boldsymbol{S}=\boldsymbol{S}_1\times\boldsymbol{S}_2\times\cdots\times\boldsymbol{S}_{n-1}\times\boldsymbol{S}_n$. $\boldsymbol{s}$ 标记为包含参与博弈的每个参与者的个人策略组合，即 $\boldsymbol{s}=\{s_1,s_2,\cdots,s_{n-1},s_n\}\in\boldsymbol{S}$. 则除去参与者 i 外的其他 $n-1$ 个博弈者的个人策略组合表示为 $\boldsymbol{s}_{-i}=\{s_1,s_2,\cdots,s_{i-1},s_{i+1}\cdots,s_{n-1},s_n\}$. 记 $\boldsymbol{s}_{-i}$ 的策略组合数为 m_{-i}，则 $m_{-i}=\prod_{k=1,k\neq i}^{n}m_k$. 对于参与者 i 的具有标量值的报酬函数为 $\pi_i(s_i,\boldsymbol{s}_{-i})$. 考虑时间维度，参与者 i 在时间 t 时段的策略则表示为 $s_i(t)$，其他参与者的策略集合(向量)则为 $\boldsymbol{s}_{-i}(t)$，相应地记参与者 i 的报酬 $\pi_i[s_i(t),\boldsymbol{s}_{-i}(t)]$.

EWA 学习算法假定每个策略都有一个吸引值. 该模型定义吸引值的初始值，吸引值如何根据经验进行更新，以及根据吸引值确定选择该策略的概率. 算法的核心是每轮更新的 2 个变量. 第 1 个变量是吸引值 $A_i^j(t)$，定义为选择策略 j 后参与者 i 的吸引值. 吸引值的更新规则为当前吸引值 $A_i^j(t)$ 是含衰减系数 φ 的上一时间吸引值 $A_i^j(t-1)$ 与当前(虚拟)报酬 $\pi_i[s_i^j,\boldsymbol{s}_{-i}(t)]$ 之和，并由更新的经验权重 $N(t)$ 归一化.

$$A_i^j(t)=\frac{\varphi\cdot N(t-1)\cdot A_i^j(t-1)}{N(t)}+\frac{\{\delta+(1-\delta)\cdot I[s_i^j,s_i(t)]\}\cdot\pi_i[s_i^j,\boldsymbol{s}_{-i}(t)]}{N(t)}. \tag{1}$$

其中：$I[\cdot]$ 指示函数为

$$I(x,y)=\begin{cases}1, & x=y;\\ 0, & x\neq y.\end{cases} \tag{2}$$

该算法在每轮博弈后用选择该策略后获得的实际报酬和模拟本轮博弈选择其他备选策略后所获得的虚拟报酬乘以权重系数 δ 对吸引值进行更新. 参数 $\delta(0<\delta<1)$ 在吸引值更新中用以衡量相对于实际获得报酬的虚拟报酬的权重. 虚拟报酬可以被理解为对选择其他备选策略所获得报酬的估计和预期.

第 2 个变量是 $N(t)$，定义为过去阅历经验折合成当前经验的观测等价值. $N(t)$ 值越大，过去阅历经验对当前吸引值的影响越大. 经验权值 $N(t)$ 的更新规则是乘以衰减系数 $\rho(0<\rho<1)$ 上一时期的经验权值 $N(t-1)$ 加上当前经验增量值 1.

$$N(t)=\rho\cdot N(t-1)+1. \tag{3}$$

定义 $A_i^j(0)$ 和 $N(0)$ 为变量 $A_i^j(t)$ 和 $N(t)$ 的初始值. 其中 $N(0)$ 可理解为博弈前思考对吸引值影响的等价评估值.

吸引值确定选择相应策略的概率. 即概率 $P_i^j(t)$ 随着吸引值 $A_i^j(t)$ 单调递增，而随 $A_i^l(t)$(其中 $l\neq j$)单调递减. 采用指数形式的选择概率表达式为

$$P_i^j(t+1)=\frac{e^{\lambda\cdot A_i^j(t)}}{\sum_{k=1}^{m_i}e^{\lambda\cdot A_i^k(t)}}. \tag{4}$$

参数 λ 用于衡量参与者对吸引值的敏感程度. 参与者对吸引值越敏感，其值也越大.

2 信道选择模型

在无线电通信信道选择问题上，不同的无线信道应具有不同的信道可用性，即认知无线电所感知的不同信道的可用概率 α 应有区别，设该无线电传输环境可划分为 n 个信道，则信道 $i(1\leqslant i\leqslant n)$ 的可用概率可用 α_i 表示，写成向量的形式 $\boldsymbol{A}=\{\alpha_1,\alpha_2,\cdots,\alpha_{n-1},\alpha_n\}$. 认知无线电在信道 i 传输数据成功完成的概率可用 β_i 表示，写成向量的形式 $\boldsymbol{B}=\{\beta_1,\beta_2,\cdots,\beta_{n-1},\beta_n\}$. 考虑无线信道特性随着时间的变化，不同时刻的信道空闲概率和信道完成概率亦不同. 则上述概率引入时间参数 t 后的形式分别为 $\boldsymbol{A}(t)=\{\alpha_1(t),\alpha_2(t),\cdots,\alpha_{n-1}(t),\alpha_n(t)\}$ 和 $\boldsymbol{B}(t)=\{\beta_1(t),\beta_2(t),\cdots,\beta_{n-1}(t),\beta_n(t)\}$.

根据 EWA 经验权值 $N(t)$ 的更新为式(3).

而吸引值 $A_i^j(t)$ 的更新表达式则应做相应的修改，以适应认知无线电信道选择和传输特点. 改进后的数学表达式为

$$A_i^j(t)=\frac{\varphi\cdot N(t-1)\cdot A_i^j(t-1)}{N(t)}+\{\delta+(1-\delta)\cdot I[s_i^j,s_i(t)]\}\cdot\pi_i[s_i^j,\boldsymbol{s}_{-i}(t)]\cdot(\eta+(1-\eta)\cdot I[1,x(j)]\cdot I[s_i^j,s_i(t)]+(1-\eta)\cdot\{1-I[s_i^j,s_i(t)]\})N(t). \tag{5}$$

其中：

$$x(j)=\begin{cases}0,\text{策略选择信道}j\text{作为优先信道通信传输失败;}\\1,\text{策略选择信道}j\text{作为优先信道通信传输成功.}\end{cases} \tag{6}$$

$$\pi_i[s_i^j,\boldsymbol{s}_{-i}(t)]=\begin{cases}0, & \text{信道}j\text{感知繁忙不可用;}\\1, & \text{信道}j\text{感知空闲可利用.}\end{cases} \tag{7}$$

则信道选择概率 $P_i^j(t+1)$ 由式(8)决定：

$$P_i^j(t+1)=\frac{e^{\lambda\cdot A_i^j(t)}}{\sum_{k=1}^{n}e^{\lambda\cdot A_i^k(t)}}. \tag{8}$$

由式(7)可知，在认知无线电进行无线环境传输信道感知搜索阶段，当感知到信道 j 当前状态繁忙(干扰严重以致无法选择该信道进行传输)时，将信道 j 的可用状态标记为 0，即不可用. 则选择信道 j 策略应得不到任何报酬，即其选择信道 j 策略相应的报酬函数 $\pi_i[s_i^j,\boldsymbol{s}_{-i}(t)]$ 取值为 0；感知到信道 j 当前状态空闲(干扰低于传输门限可选择该信道进行传输)时，将信道 j 的可用状态标记为 1，即可用. 则选择信道 j 策略理应得到相应的报酬，设此时的报酬函数 $\pi_i[s_i^j,\boldsymbol{s}_{-i}(t)]$ 值为 1.

当认知无线电进行无线环境传输信道感知搜索结束后，状态繁忙的信道所得的报酬为 0，而状态空闲的信道将标记为可用候选信道，当前状态可用但未被策略选择的信道可获得 $\delta\cdot\pi_i[s_i^j,\boldsymbol{s}_{-i}(t)]$ 的(虚拟)报酬. 本文创新点在于在式(5)引入了信道感知反馈机制. 将当前状态可用且实际被策略选中的信道的吸引值分两个阶段进行更新. 第一阶段当前状态可用且被策略选中的信道在传输开始前即获得 $\eta\cdot\pi_i[s_i^j,\boldsymbol{s}_{-i}(t)]$ 的报酬，其中 $0<\eta<1$；第二阶段当选择的信道 j 成功完成传输后将获得 $(1-\eta)\cdot\pi_i[s_i^j,\boldsymbol{s}_{-i}(t)]$ 的报酬，而当选择的信道 j 在传输过程中通信中断失败则得不到相应的反馈报酬.

3　仿真分析

假设模拟仿真环境下的无线电传输信道数为 5，即 $n=5$. 初始状态下各信道的选择概率相同 $P_i^j(0)=1/n=1/5=0.2$. 经验权值归一化的初值 $N(0)=1$，EWA 吸引值归一化的初值 $A_i^j(0)=1$. 衰减系数和敏感系数根据普遍一般经验选取其默认值 $\varphi=\rho=\lambda=0.9$. 因传输成功所获得的归一化报酬为 1，传输失败所获得的报酬为 0，假定在无任何先验概率的条件下，传输成功和失败的概率相同，则虚拟报酬权重系数 $\delta=1\times0.5+0\times0.5=0.5$. 由上节分析可知，反馈报酬权重系数应比虚拟报酬权重系数小，本文取其值一半 $\eta=\delta/2=0.25$. 为体现信道可用概率的一般性，本文选取在 0 至 1 范围内均匀分布的概率向量作为各信道可用概率，即初始信道可用概率向量 $\boldsymbol{A}_0=\{0.4,0.9,0.6,0.5,0.7\}$，则相应的信道数据传输成功完成概率向量 $\boldsymbol{B}_0=\{3/4,8/9,5/6,4/5,6/7\}$，则信道的可用概率 $\boldsymbol{\Gamma}_0=\boldsymbol{A}_0\cdot\boldsymbol{B}_0=\{0.3,0.8,0.5,0.4,0.6\}$. 为验证信道在可用概率发生改变的情况下，该智能算法能否在线实时准确决策和引导认知无线电切换到新的可用概率最高的传输信道的有效性，在整个模拟仿真过程中(36 回合后)，初始的信道空闲概率向量 $\boldsymbol{A}_0$ 将改变为 $\boldsymbol{A}_1=\{0.6,0.4,0.7,0.9,0.5\}$，相应的信道数据传输成功完成概率向量 $\boldsymbol{B}_1=\{5/6,3/4,6/7,8/9,4/5\}$，则信道的可用概率为 $\boldsymbol{\Gamma}_1=\boldsymbol{A}_1\cdot\boldsymbol{B}_1=\{0.5,0.3,0.6,0.8,0.4\}$.

本文采用简单多次重复概率实验的方法验证基于 EWA 学习的信道选择算法的有效性. 即采用回合制，每轮产生一个独立的[0,1]范围内均匀分布的随机数，如小于信道可用概率 α_i，则判断信道 i 为空闲可用状态，否则判断该信道为繁忙不可用状态. 在确定优选信道后产生一个独立的[0,1]范围内均匀分布的随机数，如小于传输数据成功完成的概率 β_j，则判断在优选信道 j 通信传输成功，否则判断在该优选信道上通信传输失败.

以上参数设定后，基于 EWA 学习的信道选择概率见图 1.

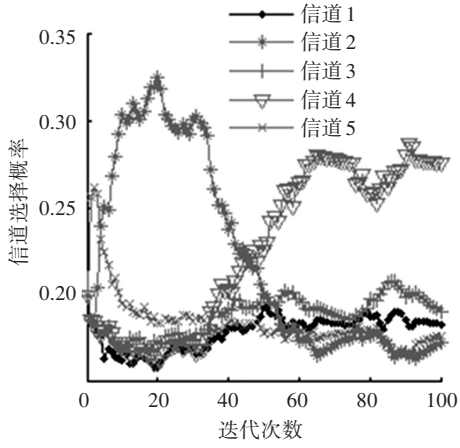

图 1　基于 EWA 学习的信道选择概率轨迹

由图 1 可见，基于 EWA 学习的信道选择算法，在初始各信道等概率的情况下随机地选择信

道 5 作为接入信道,在初始化的过渡过程后,成功地判断跟踪和锁定信道 2 作为其优选信道,其选择概率在 0.3 附近小幅波动. 在第 36 回合后,由于可用概率的变化,信道 2 的选择概率急剧下降,而信道 4 的选择概率迅速攀升,并在第 45 回合后稳定超过信道 2 的选择概率,信道 4 最终取代信道 2 成为新的信道可用概率状态下的优选信道,其选择概率最终在 0.27 附近达到稳定状态.

相应的基于 EWA 学习的信道吸引值曲线见图 2. 对比图 1,可发现吸引值的变化趋势与信道选择概率的变化趋势具有相对一致的特性.

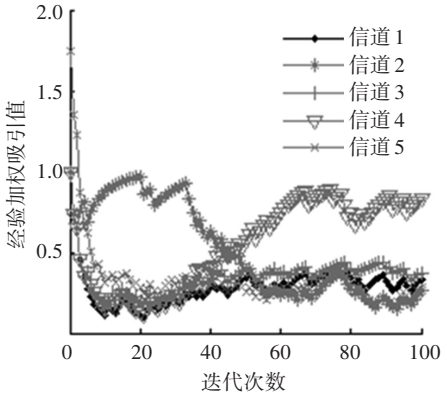

图 2 基于 EWA 学习的信道吸引值轨迹图

为便于凸显引入的 EWA 学习算法相对于普通通信道选择策略以及基于 Q 学习的信道选择策略的优越性. 现将 100 回合后的成功接入传输信道以及成功完成传输数与一直选择信道 2、信道 4 作为传输信道和基于 Q 学习的信道选择策略作统计对比. 对比结果见图 3.

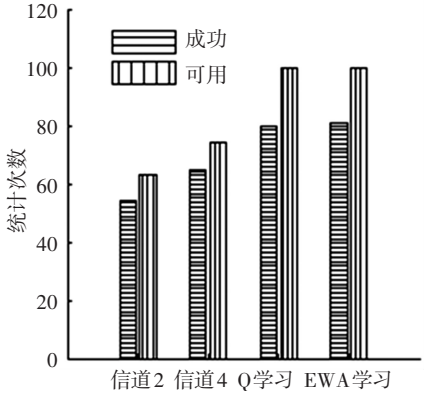

图 3 参考策略与 EWA 学习策略的对比

选择信道 2 作为通信传输信道成功接入数为 63,成功完成通信传输数 54;选择信道 4 作为通信传输信道成功接入数为 74,成功完成通信传输数 65;基于 Q 学习的信道选择策略最终通信成功完成数是 80,而 EWA 的成功完成通信传输数为 81. 通过该对比可看出,基于 Q 学习的信道选择策略相对于传统的单一信道策略在性能表现上已有显著提升. 而基于 EWA 信道选择策略的成功完成通信传输的概率则比基于 Q 学习的信道选择策略略有提升,而其优势在下图的实时信道选择对比图中将有更直观的体现.

在相同信道环境和初始条件下基于 EWA 和 Q 学习的信道选择算法的实时信道选择对比图见图 4.

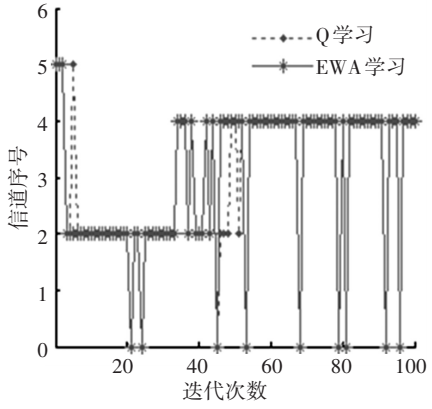

图 4 基于 EWA 学习和 Q 学习的信道选择结果

需要说明的是,当算法选择信道 0 时,定义为全信道阻塞,即此时所有的信道均处于繁忙不可用的状态,如图 4 中的第 21 回合,则表明认知无线电第一次遭遇了全信道阻塞情况. 基于 EWA 和 Q 学习的信道选择算法出现第一次差异出现在第 6 回合,Q 学习的信道选择在第 4、5 回合之后都选择第 2 信道作为优选信道的情况下,为什么 Q 学习在第 6 回合选择第 5 信道作为其传输信道. 为分析产生这一现象的原因,导出记录每回合后各信道选择概率表,见表 1、2.

表 1 基于 EWA 学习的信道选择概率表

迭代次数	信道 1	信道 2	信道 3	信道 4	信道 5
0	0.200 0	0.200 0	0.200 0	0.200 0	0.200 0
1	0.186 0	0.186 0	0.186 0	0.186 0	0.256 1
2	0.184 5	0.184 5	0.184 5	0.184 5	0.262 1
3	0.179 1	0.204 2	0.179 1	0.179 1	0.258 5
4	0.176 1	0.239 5	0.176 1	0.176 1	0.232 3
5	0.163 9	0.250 6	0.180 5	0.180 5	0.224 5
6	0.168 7	0.249 3	0.182 3	0.182 3	0.217 4

表 2 基于 Q 学习的信道选择概率表

迭代次数	信道 1	信道 2	信道 3	信道 4	信道 5
0	0.200 0	0.200 0	0.200 0	0.200 0	0.200 0
1	0.180 8	0.180 8	0.180 8	0.180 8	0.276 9
2	0.173 5	0.173 5	0.173 5	0.173 5	0.306 1
3	0.169 7	0.169 7	0.169 7	0.169 7	0.321 3
4	0.171 3	0.208 7	0.171 3	0.171 3	0.277 5
5	0.170 9	0.237 4	0.170 9	0.170 9	0.249 8
6	0.176 6	0.230 3	0.176 6	0.176 6	0.240 0

表 1 记录的基于 EWA 学习算法每回合计算得出的信道选择概率值;相应的表 2 记录的基于 Q 学习算法每回合计算得出的信道选择概率值. 由表 1 可直观地看出,第 5 回合后信道 2 的选择概率最大,其值为 0.249 3;而表 2 中,信道 5 的选择概率虽在第 4、5 回合呈现连续下降的趋势,但其值依然处于最大值 0.240 0,而此时的信道 5 在经历了第 4、5

回合的繁忙不可用状态后，在第 6 回合恢复了空闲可用状态. 所以，在第 6 回合重新选择信道 5 作为其优选信道是 Q 学习的信道选择按照其算法规则必然判断和选择的结果. 这也直接体现了基于 EWA 和 Q 学习的信道选择算法的本质差异和优势所在：基于 EWA 学习的信道选择算法具有稳定的更快跟踪锁定切换到最优信道的性能. 而在第 45 回合全信道阻塞后二者信道选择差异的原因同样可以用上述分析进行解释.

由此，基于 EWA 的信道选择算法比 Q 学习的信道选择策略具有更强的实时应变性，具有更加快速地发现和适应信道环境变化的能力.

4　结　语

本文创新地提出基于 EWA 学习的信道选择算法，引入了信道传输反馈机制，使得系统能够在线学习井下无线电环境通信信道特性，通过积累自我历史选择信道经验，推测并选择和更改当前最优通信信道，实时动态保证通信链路质量，降低系统通信中断概率. 通过模拟仿真程序简单多次重复概率实验的方法验证该智能算法能够在线实时准确决策并引导认知无线电切换到新的可用概率最高的传输信道，与传统选择单一通信道和 Q 学习的信道选择策略相比，基于 EWA 学习算法的认知无线电具有更加稳定的性能和更加快速准确捕获优选通信传输信道的能力.

参考文献

[1] 胡穗延.煤矿自动化和通信技术现状与发展趋势[J].煤炭科学技术，2007(8)：1-4.

[2] 赵红玉，陶晋宜，王园媛，等.矿井无线透地通信中语音信号压缩的研究[J].煤炭工程，2014(3)：20-22.

[3] 张国鹏，王艳芬，丁恩杰.矿井无线多媒体传感器网络 UWB 信号收发策略研究[J].煤炭科学技术，2013(12)：71-75.

[4] 郭银景，杨洪骞，杨涛，等.一种井下无线自组织网络自愈路由算法[J].山东科技大学学报：自然科学版，2012(1)：84-88.

[5] 张婧.基于 Ad Hoc 的矿井巷道无线传感器网络路由协议分析[J].山西大学学报：自然科学版，2013(4)：586-590.

[6] 陈捷，阚保强.衰落信道下的无线传感器网络虚拟 MIMO 传输优化设计[J].山东科技大学学报：自然科学版，2013(5)：96-101.

[7] 王志刚.一种无线 Mesh 网络煤矿监控系统实现[J].煤炭技术，2014(1)：161-163.

[8] 盛慧兴，方志远，霍冠英.OFDM 技术及在认知无线电中应用[J].信息安全与保密，2007(5)：71-73.

[9] 王文清.煤矿井下无线电噪声的研究[J].煤炭学报，2011(10)：1773-1776.

[10] 张守祥，王汝琳，刘芳.基于 OFDM 的巷道和工作面无线移动通信研究[J].煤矿安全，2006(3)：13-16.

[11] 孙继平，成凌飞，张长森.截面尺寸对矩形巷道中电磁波传播的影响[J].中国矿业大学学报，2005(9)：596-599.

[12] 孙继平，成凌飞，张长森.电导率对巷道中电波传播的影响[J].辽宁工程技术大学学报，2007(2)：96-98.

[13] MITOLA J. Cognitive radio for flexible mobile multimedia communications [J]. Mobile Networks & Applications, 2001, 6(5)：435-441.

[14] BANTOUNA A, STAVROULAKI V, KRITIKOU Y, et al. An overview of learning mechanisms for cognitive systems [J]. Eurasip Journal on Wireless Communications and Networking, 2012(22)：1-6.

[15] TSAGKARIS K, KATIDIOTIS A, DEMESTICHAS P. Neural network-based learning schemes for cognitive radio systems[J]. Computer Communications, 2008, 31(14)：3394-3404.

[16] KATIDIOTIS A, TSAGKARIS K, DEMESTICHAS P. Performance evaluation of artificial neural network-based learning schemes for cognitive radio systems [J]. Computers & Electrical Engineering, 2010, 36(3)：518-535.

[17] OKSANEN J, LUNDEN J, KOIVUNEN V. Reinforcement learning based sensing policy optimization for energy efficient cognitive radio networks [J]. Neurocomputing, 2012, 80：102-110.

[18] CHEN X F, ZHAO Z F, ZHANG H G, et al. Reinforcement learning enhanced iterative power allocation in stochastic cognitive wireless mesh networks[J]. Wireless Personal Communications, 2011, 57(1)：89-104.

[19] TENG Y L, YU F R, HAN K, et al. Reinforcement-learning-based double auction design for dynamic spectrum access in cognitive radio networks[J]. Wireless Personal Communications, 2013, 69(2)：771-791.

[20] ZHANG W Z, LIU X C. Centralized dynamic spectrum allocation in cognitive radio networks based on fuzzy logic and Q-learning[J]. China Communications, 2011, 8(7)：46-54.

[21] GALINDO-RRANO A, GIUPPONI L. Distributed Q-learning for aggregated interference control in cognitive radio networks [J]. Ieee Transactions on Vehicular Technology, 2010, 59(4)：1823-1834.

[22] ZHU J, WANG J, LUO T, et al. Adaptive transmission scheduling over fading channels for energy-efficient cognitive radio networks by reinforcement learning[J]. Telecommunication Systems, 2009, 42(12)：123-138.

（编辑　苗秀芝）