doi:10.3969/j.issn.1673-9833.2013.02.016

一种基于模糊Q学习算法的认知无线电频谱分配策略

徐 勇，果 鑫，刘丰年，文 鸿，张文平，李长云

（湖南工业大学 计算机与通信学院，湖南 株洲 412007）

摘 要：认知无线电是一种智能推理学习的通信系统，为了实现认知无线电频谱分配智能学习过程，引入模糊Q学习方法。认知用户通过在线Q学习来调节模糊推理系统，得到最优的频谱分配模糊规则，实现自适应的频谱分配方案。最后将模糊Q频谱分配算法与非智能学习算法（模糊频谱分配算法以及随机分配算法）进行比较，仿真结果证明了该方案能在一定程度上提高系统带宽收益，同时降低系统的冲突率。

关键词：认知无线电；自适应；模糊Q；频谱分配；带宽收益；冲突率

中图分类号：TN915.9 **文献标志码**：A **文章编号**：1673-9833(2013)02-0074-05

A Spectrum Allocation Strategy Based on Fuzzy Q Learning Algorithm for Cognitive Radio

Xu Yong，Guo Xin，Liu Fengnian，Wen Hong，Zhang Wenping，Li Changyun

(School of Computer and Communications, Hunan University of Technology, Zhuzhou Hunan 412007, China)

Abstract: Cognitive radio is a kind of intelligent reasoning learning communication system, and fuzzy Q learning method is introduced to realize the cognitive radio spectrum allocation intelligent learning process. Cognitive users adjust the fuzzy inference system through online Q learning, obtain the optimal fuzzy rules of spectrum allocation and realize the adaptive spectrum allocation scheme. The fuzzy Q spectrum allocation algorithm is compared with non intelligent learning algorithm (fuzzy spectrum allocation algorithm and random distribution algorithm). And the simulation results show that this algorithm improves the system bandwidth income and reduces the system conflict rate.

Keywords: cognitive radio; adaptive; fuzzy Q; spectrum allocation; bandwidth income; conflict rate

0 引言

为了解决日益增加的用户需求与匮乏的频谱资源之间的矛盾，人们提出了认知无线电（cognitive radio）技术。认知无线电是一种提高有限频谱资源利用率的新形式，它能够感知外界环境，并能使用人工智能技术从环境中学习[1]，在不影响授权用户的前提下，认知用户可以利用授权用户的空闲频谱实现可靠通信。

频谱分配是认知无线电的关键技术之一，根据接入方式的不同可以分为填充式（underlay）和下垫式（overlay）2种分配方式。对于频谱分配策略问题，人们进行了大量的研究，如Zheng H.等人提出了基于图论着色模型的频谱接入方案[2]；D. Niyato等人提出了基于博弈论的频谱分配模型[3]；Zhang Yonghong等人提出了基于正交频分复用技术（orthogond frequency

收稿日期：2012-12-15

基金项目：湖南省研究生科研创新基金资助项目（CX2011B393），湖南省自然科学基金资助项目（11JJ3002）

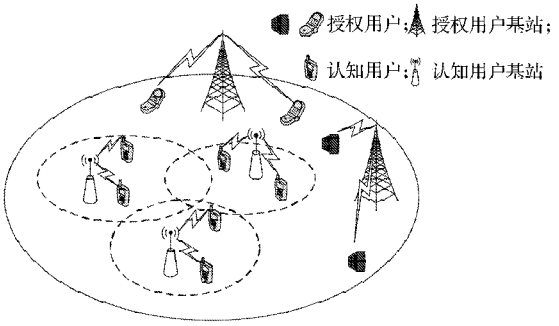
作者简介：徐 勇，（1988-），男，湖北荆州人，湖南工业大学硕士生，主要研究方向为认知无线电频谱分配，

E-mail：xy323@126.com

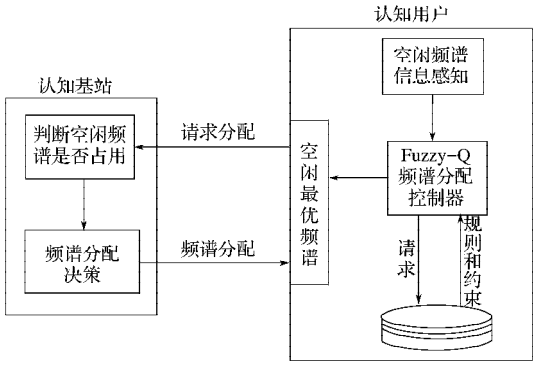
division multiplexing, OFDM）的频谱分配方案[4]，以及赵青等人提出的基于 OFDM 的新算法[5]中，认知用户发射功率与授权用户干扰温度受限的条件下，认知用户尽可能长时间地使用授权用户的频谱信道。上述各方案中，频谱分配过程都不偏向认知用户对环境的学习自适应能力。本文提出了一种基于模糊 Q 学习的频谱分配方案，将 Q 学习引入认知用户频谱分配问题中，使认知用户具备一定的学习推理能力，在保证系统总带宽收益和认知用户的公平接入的同时，认知用户能自动感知环境状态，实现自适应的频谱分配。

1　认知无线电 Fuzzy Q 频谱分配模型

本文主要讨论集中式认知无线电频谱分配模型，如图 1 所示，一个小区内分布着若干个授权用户和认知用户，授权用户主要通过授权用户基站进行通信，而认知用户则依赖认知基站进行相关通信。授权用户出现的概率是随机的，授权用户是否通信的概率也是随机的，认知用户根据认知基站分配的频谱资源动态的占用授权用户的频带进行通信。



a）认知无线电频谱分配网络拓扑



b）频谱分配模型

图 1　认知无线电频谱分配网络拓扑及框架

Fig. 1　The network topology and framework of cognitive radio spectrum allocation

在当前系统中，认知用户对当前环境的空闲频谱信息进行感知，作为 Fuzzy Q 控制器的输入端，经过模糊推理，以及 Q 值的学习，得到最优的感知频谱。认知基站接受各认知用户的的频谱信息，查看频谱是否被占用，未占用则分配给各认知用户，保证认知用户获得最优的频谱资源。

2　模糊 Q 学习算法

模糊 Q 学习[6]是基于模糊控制系统来实现的，用模糊分割的状态空间和离散的动作空间来实现 Q 学习，逼近最优控制策略和 Q 函数。

Q 学习是强化学习的一种方法，强化学习中的学习者称为 Agent。在 Q 学习中，Agent 不去估计环境模型，而是直接优化一个可迭代计算的 Q 函数，定义此 Q 函数为在状态 s_t 时执行动作 a_t，且此后按最优动作序列执行时的折扣累计强化值，即

$$Q_{t+1}(s_t,a_t)=r_t+\gamma\max_{a_t}Q(s_{t+1},a_t), \tag{1}$$

式中：γ 为折扣因子，且 $0\leqslant\gamma\leqslant1$；$r_t$ 为 t 时刻的回报值。

在 Q 学习中，Agent 的学习过程为：观察现在的状态 s_t，选择并执行一个动作 a_t，然后观察下一个状态 s_{t+1}，收到一个立即回报 r_t，然后根据式（2）修改 Q 值

$$Q_t(s_t,a_t)=(1-\alpha)Q_{t-1}(s_t,a_t)+\alpha\left[r_t+\gamma\max_{a_t}Q(s_{t+1},a_t)\right], \tag{2}$$

式中 α 为学习速率。

模糊控制系统能够映射一个状态 s_t 到一个动作 a_t。它同样能够在一个连续的状态空间中映射一个状态——动作对 (s_t,a_t) 到一个状态值 $Q_t(s_t,a_t)$。在模糊 Q 学习中，用 n 维向量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$ 来描述模糊 Q 学习中的状态空间，然后使用 m 个独立的值 $w_k(k=1,2,\cdots,m)$ 来决定一个连续的动作。模糊 Q 学习中模糊规则[7]可以表示为

R_j: If x_1 is A_{j1} and x_2 is A_{j2} and$\cdots$and x_n is A_{jn} then $w_j=(w_{j1},w_{j2},\cdots,w_{jm}),j=1,2,\cdots,N$;

其中，R_j 是规则集合第 j 条规则；A_{ji} 是规则 R_j 中的输入变量 x_i 的模糊集；w_j 代表模糊规则 R_j 的后件输出结果；N 代表模糊规则的数目。

Agent 在感知当前状态后，通过模糊控制系统中模糊规则推理可以输出响应的控制动作，每执行完一个动作后，Q 值会更新，使用反向传播算法调节模糊 Q 学习网络权值使误差 ΔQ 尽可能小，这样就可以实现 Q 值的学习。学习公式如式（3）所示，即

$$\Delta Q_t=\alpha\left[r_t+\gamma\max_{a_t}Q(s_{t+1},a_t)-Q(s_t,a_t)\right]。 \tag{3}$$

3　模糊 RBF 神经网络 Q 学习在频谱分配的应用

为了实现认知用户在一个连续状态空间的 Q 学

习，采用一种基于模糊RBF神经网络Q学习策略[8-10]来建立模糊推理系统，实现模糊Q学习算法。系统共有5层。

1）输入层。输入为连续的状态向量$\boldsymbol{x}_t=(x_1,x_2,\cdots,x_n)$，该层的每个节点$i$的输入输出表示为$f_1(i)=x_i$。

2）模糊化层。该层的每个节点执行一个隶属度函数功能求取各输入变量的隶属度μ_{ij}，隶属度函数采用高斯型函数，c_{ij}和σ_{ij}分别是第i个输入变量第j个模糊集合的隶属度函数的均值和标准差。即

$$f_2(i,j)=\mu_{ij}=\exp\left\{-\frac{\left(x_i-c_{ij}\right)^2}{\left(\sigma_{ij}\right)^2}\right\}。\quad(4)$$

3）规则层。该层通过与模糊化层的连接来完成模糊规则的匹配，每个节点的输出为该节点所有输入信号的乘积，即

$$f_3(j)=\prod_{j=1}^{N}f_2(i,j)=\mu_{i1}\mu_{i2}\cdots\mu_{iN},\ j=1,2,\cdots,N;\quad(5)$$

式中，$N=\prod_{i=1}^{n}N_i$，N_i是输入层中第i个输入模糊化层节点数。

4）去模糊化层。该层输出为当前状态下Agent每个动作的对应Q值，

$$f_4(l)=Q(s_t,a_t^l)=\sum_{j=1}^{N}w_{lj}\cdot f_3(j)\cdot q(s_t,a_t^l),\quad(6)$$

式中：l为去模糊化层节点个数，对应于动作对中动作的个数；$Q(s_t, a_t^l)$为s_t状态下第l个动作对应的Q值。

5）输出层。该层包含2个节点，包括最终对应的Q值和动作。去模糊化层得到每个动作对应的Q值，根据Boltzmann动作选择策略得到最终的全局动作和对应的Q值。每个动作被选择到的概率为

$$prob(a^l)=\frac{e^{Q(s,a^l)/T}}{\sum_{a_k\in A}e^{Q(s,a_k)/T}},\quad(7)$$

式中：$Q(s, a^l)$是动作对中第l个动作的Q值；T为温度系数，T越大，随机性越大；

Boltzmann动作选择算法如下：

1）根据可执行动作集$A=\{a_1,a_2,a_3,\cdots\}$中每个动作a_i对应的Q值，由式（7）得到每个动作被选择到的概率$prob(a_i)$；

2）在$[0,1]$之间生成一个随机数$temprand$；令$count=0$，$i=1$；

3）$count\leftarrow count+prob(a_i)$；

4）如果$temprand\leqslant count$，则a_i为选中动作，否则转到第3步，且$i\leftarrow i+1$。

3.1 模糊RBF神经网络Q学习中参数学习

令TD误差ΔQ为模糊RBF网络误差函数，则

$$e_{t+1}=r_{t+1}+\gamma\max_{a_t}Q(s_{t+1},a_t)-Q(s_t,a_t)。\quad(8)$$

采用梯度下降法来修正可调参数，定义目标误差函数为

$$E(t)=\frac{1}{2}e_{t+1}^2。\quad(9)$$

网络权值通过如下公式来调整：

$$\Delta w=-\eta\frac{\partial E}{\partial w}=-\eta\frac{\partial E}{\partial e_{t+1}}\frac{\partial e_{t+1}}{\partial Q_t}\frac{\partial Q_t}{\partial w_t^j}=\eta e_{t+1}f_3(j)。\quad(10)$$

式中η为学习速率，$\eta\in[0,1]$。

网络权值的学习算法为

$$w_{t+1}^j(a_{t+1}^j)=w_t^j(a_t^j)+\Delta w,\quad(11)$$

隶属度函数参数通过如下方式调整：

$$c_{ij}(t+1)=c_{ij}(t)-\eta e_{t+1}wf_3(j)\left[\frac{x_i-c_{ij}}{\sigma_{ij}^2}\right];\quad(12)$$

$$\sigma_{ij}(t+1)=\sigma_{ij}(t)-\eta e_{t+1}wf_3(j)\left[\frac{(x_i-c_{ij})^2}{\sigma_{ij}^3}\right]。\quad(13)$$

3.2 Fuzzy-Q频谱分配算法流程

Step 1：初始化Q学习及RBF神经网络参数，网络参数取随机值；

Step 2：根据给定t时刻的网络输入状态，由式（5），（6）得到动作集中每个动作对应的Q值$Q(s_t,a_t^l)$；

Step 3：使用Boltzmann动作选择策略，由式（7）得到每个动作被选择的概率，并与随机值进行比较，得到当前最优动作a_t及对应的Q值$Q(s_t,a_t)$；

Step 4：执行动作a_t，获得对应的奖惩r_t，Agent状态转移到s_{t+1}；

Step 5：当前频段被占用，保留网络权值，程序跳转到Step 2，否则继续运行；

Step 6：根据参数学习方法，更新网络权值w_t^j及对应隶属度函数的中心c_{ij}和宽度σ_{ij}；

Step 7：判断是否满足最大训练数，若满足，结束程序，否则，转到Step 2。

4 仿真与分析

4.1 频谱分配系统状态空间

状态空间是一个系统全部可能状态的集合，可以用状态特征矢量来描述。状态特征向量指包括对于智能体决策必要的有用信息，能够唯一对应地表示状态空间的特征。对于认知无线电频谱分配问题，可以选择认知用户移动速率（mobility），认知基站与认知用户之间的距离（distance）2个变量作为系统输入状态向量[11-12]。则输入状态向量表示为

$$\boldsymbol{s}_t=[Mobility, Distance]^{\mathrm{T}}。$$

对于一个空闲频段，不同认知用户检测的SNR值可能不同。频段信噪比越高，通信质量越好，通

信系统吞吐量越大，该空闲频段被分配的概率越大。根据香农定理，信道带宽的值为

$$c=B\log_2(1+SNR), \tag{14}$$

则系统总的带宽为

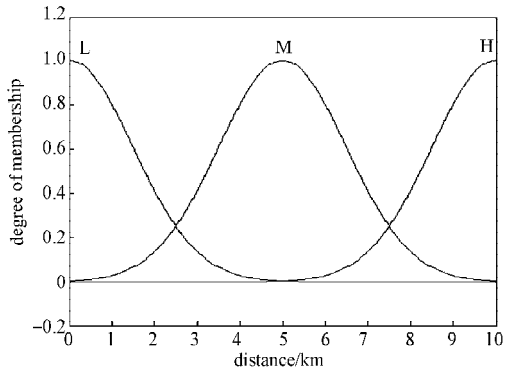
$$C=\sum_{i=1}^{m}c(i), \tag{15}$$

式中 m 为可用频段的数量。

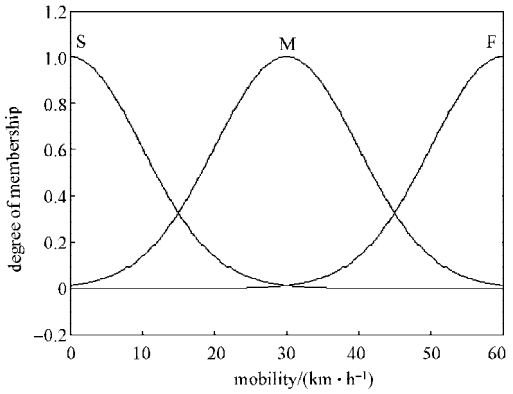
认知用户的移动速率在频谱分配中起很重要的作用，当认知用户以一定速率移动的时候会产生多普勒效应。认知用户的移动也减小了对授权用户信号检测的能力，可能导致认知用户错误地认为当前频段为空闲频段，从而产生对授权用户的有害干扰。

此外，在频谱分配中认知基站与认知用户之间的距离也是一个很重要参数。认知用户在选择好分配信道后，向基站请求分配。此时距离基站近的认知用户信道条件好，能更好地保证通信的质量。

对上述 2 个状态变量进行模糊划分，各自对应的语言值为：*Mobility*:{Slow, Moderate, Fast}，*Distance*:{Near, Moderate, Far}。图 2 为对应变量的隶属度函数。



a）距离隶属度函数



b）移动速率隶属度函数

图 2　系统隶属度函数

Fig. 2　System membership function

在频谱分配系统中，主要的目的是得到分配的最优频谱，则系统的状态输出动作可以定义为：将第 i 个空闲频段分配给认知用户，表示为 $a_i=\{a_1,a_2,a_3,\cdots,a_i\}$。

4.2　强化函数定义

强化信号即“奖惩信号”，在强化学习系统中，Agent 的目标被形式化为奖惩信号，Agent 的目标就是使得奖惩信号的总和最大。认知无线电中认知用户伺机占用频谱，所以频谱分配系统中 Agent 的奖惩信号可以依据在执行空闲频段分配动作时，被分配频段是否被占用来选取。强化函数定义如下，即

$$r=\begin{cases}-5, \text{分配频段被占用;}\\ 1, \text{分配频段未被占用。}\end{cases}$$

4.3　仿真实验

用 Matlab 7.0 搭建本实验的仿真平台，从频谱分配系统的总带宽收益与认知用户接入冲突概率两方面进行了性能仿真，对比了文献[12]中模糊频谱分配与随机分配 2 种方法。模糊频谱分配算法仿真参数参照文献[11−12]，Q 学习参数参照文献[10]设置，具体参数如表 1 所示。

表 1　仿真参数设置

Table 1　The simulation parameters

参数	参数取值
空闲频段数量	5
认知用户数量	5
信道带宽	200 kHz
信噪比	12~28（dB）
距离	0~10（km）
移动性	0~60（km/h）
温度系数	8
折扣因子γ	0.9
学习速率η	0.5
初始权值 w	0 ~ 1
仿真次数	10 000

在基站范围内随机产生空闲频段和认知用户，每个对应空闲频段授权用户出现的概率服从泊松分布，当空闲频段被分配且未被占用，由式（15）计算系统总带宽收益，当分配频段被占用时，占用计数加 1。认知用户学习 20 000 次，每 500 次为一个统计点，分 40 个相等的学习阶段统计当前频段冲突概率。仿真结果如图 3 所示。

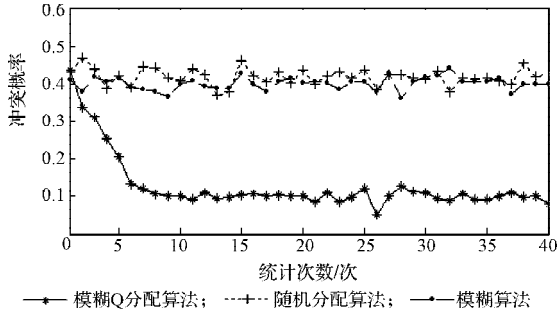


图 3　认知用户接入冲突概率

Fig. 3　The access conflict probability for cognitive users

从图3可以看出，模糊Q算法的冲突率小于模糊算法以及随机分配算法，且模糊Q算法冲突概率随迭代次数增加而递减，当迭代进行到3 000次左右，用户接入概率基本收敛。系统总的平均宽带收益如图4所示。

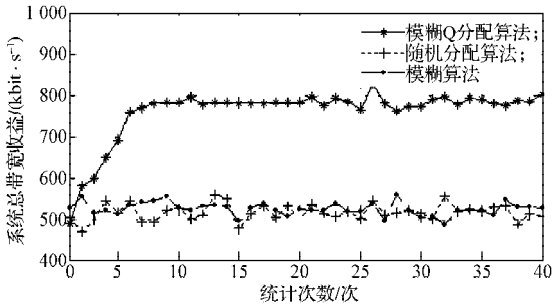


图4 系统总的平均带宽收益

Fig. 4 The average bandwidth income for overall system

从图4可以看出，模糊Q算法获得的系统总的平均带宽收益大于模糊算法以及随机分配算法，且模糊Q算法带宽收益随迭代次数的增加而增加，当迭代到3 000次左右，用户的平均带宽收益也趋于一个收敛值。

为了进一步评价算法的有效性，针对40个统计节点统计出了随迭代次数增加，每个信道被选择的概率图。仿真的5个信道中，信道占用率分别为0.8, 0.3, 0.4, 0.5, 0.1，仿真的结果如图5。

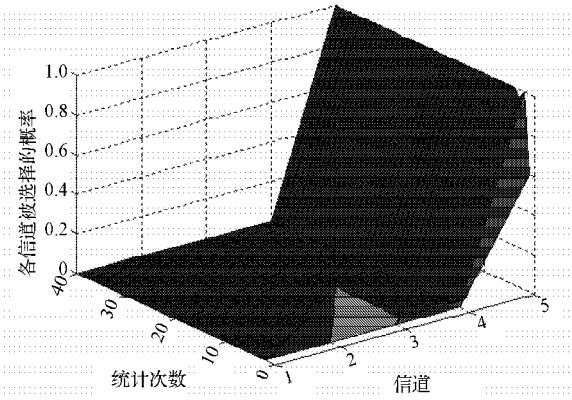


图5 信道被选择率随迭代统计次数变化图

Fig. 5 Channel selection rate varying with iterative statistical number

在第5次统计后，算法趋向收敛，算法偏向与选择最小占用率的第5信道。这显示了模糊Q算法的在线学习能力，能自适应地选择最优信道，实现频谱分配策略。

根据以上仿真结果可知，模糊Q分配算法中冲突概率随迭代次数增加而递减的过程反映了认知用户对环境的学习过程，对比模糊算法这类非学习性算法，基于模糊Q算法的频谱分配算法与环境交互的自学习能力，在一定程度上减小了系统的冲突率，增加了系统带宽收益。由于仿真中授权用户频谱可用率随机性，以及模糊Q算法中学习参数，以及温度系数选择的问题，导致仿真的结果不能每一次都最优，但是却保证仿真过程平均值最优，即模糊Q算法在上述3种算法中整体性能最优。

5 结语

本文介绍了基于模糊Q学习的频谱分配算法，通过Q学习改进模糊规则库，实现自适应的频谱分配。模糊Q算法在一定程度上减小了系统的冲突率，增大了系统带宽收益。未来研究的工作一方面优化Q学习算法，另一方面可以增加业务分类，用户偏好等状态来选择频谱信道，减小系统冲突率，提高系统带宽收益。此外，本文讨论的主要是基于overlay接入方式的频谱分配，对于在underlay接入方式下，认知用户与授权用户共存情况下的频谱分配策略没有讨论，可以在这一方面做进一步的研究。

参考文献：

[1] Federal Communicafions Commission. Spectrum Policy Task Force Report [R]. Washington DC：FCC，2002：1-10.

[2] Zheng H，Peng C. Collaboration and Fairness in Opportunistic Spectrum Access[C]// 2005 IEEE International Conference on Communications. [S.l.]: Conference Publications，2005：3132-3136.

[3] Niyato D，Hossain E. A Game-Theoretic Approach to Competitive Spectrum Sharing in Cognitive Radio Networks [C]//in IEEE Wireless Communications and Networking Conference. Kowloon：Conference Publications，2007：16-20.

[4] Zhang Yonghong，Cyril Leung. Resource Allocation in an OFDM-Based Cognitive Radio System[J]. IEEE Transaction on Communications，2009，57(7)：1928-1931.

[5] 赵 青，朱 琦. OFDM认知无线电系统中多用户资源分配新算法[J]. 电路与系统学报，2011，16(4)：44-50.
Zhao Qing，Zhu Qi. A Novel Resource Allocation Algorithm for Multiuser OFDM-Based Cognitive Radio Systems[J]. Journal of Circuits and Systems，2011，16(4)：44-50.

[6] 张汝波.强化学习理论及应用[M].哈尔滨：哈尔滨工程大学出版社，2001：126-139.
Zhang Rubo. Reinforcement Learning Theory and Applications[M]. Harbin：Harbin Engineering University Press，2000：126-139.

[7] Ishibuchi H，Nakashima T，Miyamoto H，et al. Fuzzy Q-Learning for a Multi-Player Noncooperative Repeated

（下转第88页）