

文章编号: 1001-0920(2013)11-1643-07

机会式频谱接入问题基于事件的优化方法

黄永皓, 陈曦

(清华大学 a. 智能技术与系统国家重点实验室, b. 智能与网络化系统研究中心, 北京 100084)

摘要: 研究机会式频谱接入技术中次用户对可利用频谱进行探测和接入策略的优化问题. 通过引入事件的概念, 将含有可数无限状态的原问题转化为包含有限个事件的决策问题. 从性能灵敏度的角度出发, 分析不同策略下平均传输率的差异, 给出了基于事件策略的性能差分公式. 以此为基础, 通过合理的近似, 设计了基于事件的策略迭代算法. 仿真示例验证了所提出算法的有效性和近似处理的合理性.

关键词: 机会式频谱接入; 马尔可夫决策过程; 基于事件的优化; 灵敏度分析; 策略迭代

中图分类号: TP13

文献标志码: A

Event-based optimization approach to opportunistic spectrum access problem

HUANG Yong-hao, CHEN Xi

(a. Tsinghua National Laboratory for Information Science and Technology, b. Center for Intelligent and Networked Systems, Tsinghua University, Beijing 100084, China. Correspondent: HUANG Yong-hao, E-mail: huang-yh02@mails.tsinghua.edu.cn)

Abstract: The sense and access optimization problem in opportunistic spectrum access technology is considered. By introducing the concept of event, the original problem with countably infinite states is converted to an event-based decision problem with finite events. From a sensitivity-based view, the performance differences between two event-based policies are investigated. Based on the event-based performance difference formula, with the appropriate approximation, the sample-path based policy iteration algorithm is developed. An example is provided to illustrate the effectiveness and reasonableness of the proposed algorithm.

Key words: opportunistic spectrum access; Markov decision process; event-based optimization; sensitivity-based approach; policy iteration

0 引言

近年来, 无线通信系统的快速发展与有限的频谱资源之间的矛盾日益受到关注. 为了解决无线频谱资源使用不均衡的问题, 人们提出了动态频谱接入技术的全新通信模式. 其中, 机会式频谱接入由于与现有通信系统兼容性较好, 已成为当前动态频谱接入技术研究中的主要模型^[1].

在机会式频谱接入技术中, 通信系统中的信道被分配给某个固定的主用户, 同时允许其他次用户在不影响主用户使用的前提下使用. 本文研究次用户对信道进行探测和接入策略的优化问题. 现有研究只限于讨论基于部分可观模型下的次优策略或特定情况下

的最优策略^[2-6]. 文献[7]研究了与该问题的部分可观马尔可夫过程等价的信度马尔可夫过程, 尽管在理论上给出了最优策略的求解算法, 但在实际应用时, 需对信度马尔可夫过程的连续状态空间进行集结, 而文献[7]没有研究分析状态集结对解的最优性的影响.

本文通过引入事件的概念^[8], 将该问题转化为包含有限个事件的基于事件的决策过程问题. 因为事件的定义对应于真实系统中的物理现象, 所以这种模型比较直观. 同时事件的定义又有一定的灵活性, 可以根据实际存储或计算条件的限制决定事件的数量. 通过采用基于性能灵敏度分析的方法^[9], 本文分析了不同基于事件策略下次用户的平均传输率的差异, 并推

收稿日期: 2012-07-16; 修回日期: 2012-10-21.

基金项目: 国家自然科学基金项目(61203039); 高等学校学科创新引智计划(B06002); 清华信息科学与技术国家实验室(筹).

作者简介: 黄永皓(1983-), 男, 博士生, 从事马尔可夫决策过程的研究; 陈曦(1965-), 女, 副研究员, 从事随机最优控制、传感器网络等研究.

导出基于事件策略的性能差分公式. 在此基础上, 通过合理的近似, 开发出基于事件的策略迭代算法, 并分析了近似处理带来的误差上限. 这种近似处理的误差随着事件数量的增加而减小, 因此可以根据实际需要进行控制. 当实际所考虑的策略为有限个时, 如果误差上限小于任意两个策略的性能之差, 误差可以忽略, 策略迭代算法收敛到最优策略.

1 问题描述

在机会式频谱接入模型中, 系统包含 N 个信道, 每个信道被分配给某个固定的用户使用, 此用户被称为该信道的主用户. 在每个时隙(时间段), 根据主用户是否使用信道, 信道的状态分为空闲状态和忙碌状态, 分别用 I 和 B 表示. 假设每个信道的状态变化可以用一条离散时间马尔可夫链描述, 且所有信道的状态转移特性相同, 即每个信道的状态转移概率均为 p_{uv} , $u, v \in \{I, B\}$. 为了提高信道的利用率, 除了 N 个主用户之外, 允许其他用户在信道空闲时使用, 这些用户被称为次用户. 次用户使用信道的前提是不影响该信道主用户的通信服务. 因此, 次用户在使用信道之前, 需要先对信道状态进行探测, 只有在信道空闲时, 次用户才接入使用; 若当前时隙被探测信道处于忙碌状态, 则次用户放弃接入, 等待到下一时隙重新选择信道进行探测. 由于次用户在每个时隙只能获知一个信道的确切状态, 为了更好地作出决策, 对于每个信道, 次用户不仅要保存上一次探测所获知的信道状态, 还需要保存上一次探测的时间. 通过这些信息, 次用户可以估计出当前时隙每个信道处于空闲状态概率, 并基于此决定探测哪条信道. 这里并不要求次用户已知信道的状态转移概率 p_{uv} , 若 p_{uv} 未知, 则次用户仍可以根据观测历史估计信道状态. 次用户的目标是通过选择合适的探测与接入策略, 实现自己数据传输率的最大化. 一般情况下, 次用户的最优策略并不一定是选择当前时隙处于空闲状态概率最大的信道进行探测, 因为次用户在当前时隙的选择不仅影响当前时隙的数据传输量, 而且还影响到未来对各个信道状态的估计, 从而影响到未来的数据传输量, 所以, 次用户的最优策略既要考虑当前时隙的数据传输量, 也要保证未来的数据传输量, 从而达到长期最优.

首先, 本文以马尔可夫决策过程为模型, 采用基于灵敏度分析的方法, 简要讨论该问题的求解. 以次用户关于系统所有信道的信息作为系统状态, 包括次用户对每个信道上一次探测所获知的状态和上一次探测的时间, 记为 $S(t) = (S_1(t), S_2(t), \dots, S_N(t))$. 关于每个信道的信息, 采用以下表示方式: 在第 t 个时隙, 如果在 k 个时隙之前探测到该信道的状态为空闲 (I), 则将次用户对信道 i 的认识表示为 $S_i(t) = 2k - 1$; 反之, 如果在 k 个时隙之前探测到该信道的

状态为忙碌 (B), 则将次用户对信道 i 的认识表示为 $S_i(t) = 2k$, $k = 1, 2, \dots$; 因此, 系统的状态空间为 $\mathcal{S} = \mathbf{N}^N$, 其中 \mathbf{N} 表示正整数集. 采用这样的表示方式, 不仅可以通过 $S_i(t)$ 计算出上一次对信道 i 的探测时间, 还可以根据 $S_i(t)$ 的奇偶性判断出上一次探测信道 i 的状态.

在第 t 个时隙, 设系统状态为 $S(t) = s$, 次用户依照某既定策略 Φ 选择信道 i 进行探测. 这里本文只考虑平稳策略 $\Phi: \mathcal{S} \rightarrow \{1, 2, \dots, N\}$, 即 Φ 是从状态空间 \mathcal{S} 到信道集合 $\{1, 2, \dots, N\}$ 的映射. 在选择信道 $i = \Phi(s)$ 进行探测之后, 次用户会根据信道 i 的状态决定是否接入信道, 若信道 i 空闲, 则次用户使用该信道传输数据, 数据传输量增加 1; 若信道 i 忙碌, 则次用户不传输任何数据, 数据传输量为 0. 在策略 Φ 下, 次用户在系统状态为 $S(t) = s$ 时数据传输量的期望值记为 $R^\Phi(s)$. 为了计算信道 i 处于空闲的概率以及 $R^\Phi(s)$, 与文献 [7] 中相同, 定义算子 $\tau(\alpha) := \alpha p_{II} + (1 - \alpha)p_{BI}$, $\forall \alpha \in [0, 1]$, 则 $\tau(\alpha)$ 表示当前时刻信道处于的空闲概率为 α 时, 下一时刻仍处于空闲的概率. 同时记 $\tau^{k+1}(\alpha) = \tau(\tau^k(\alpha))$, 且 $\tau^0(\alpha) = \alpha$. 由此, 根据所探测信道 i 的状态 k_i , 可以计算其处于空闲的概率及数据传输量的期望值. 如果 k_i 是奇数, 则该信道处于空闲的概率为 $\tau^{(k_i-1)/2}(p_{II})$, 次用户选择信道 i 的数据传输量的期望值为 $R^\Phi(S(t)) = \tau^{(k_i-1)/2}(p_{II}) \cdot 1$; 同理, 如果 k_i 是偶数, 则数据传输量的期望值为 $R^\Phi(S(t)) = \tau^{(k_i-2)/2}(p_{BI}) \cdot 1$.

从第 t 个时隙到第 $t+1$ 个时隙, 根据次用户选择探测信道的不同, 系统状态发生相应地变化. 设当前时隙的系统状态为 $S(t) = (k_1, k_2, \dots, k_N)$, 如果次用户探测到信道 i 处于空闲状态 (I), 则系统状态变为

$$S(t+1) = (k_1 + 2, k_2 + 2, \dots, k_{i-1} + 2, 1, k_{i+1} + 2, \dots, k_N + 2);$$

反之, 如果次用户探测到信道 i 处于忙碌状态 (B), 则系统状态变为

$$S(t+1) = (k_1 + 2, k_2 + 2, \dots, k_{i-1} + 2, 2, k_{i+1} + 2, \dots, k_N + 2).$$

假设系统在所有平稳策略下都是遍历马尔可夫链, 并记系统在策略 Φ 下的状态转移函数记为 P^Φ . 随着时间的推移, 在策略 Φ 下, 次用户的长期平均数据传输率为

$$\eta^\Phi = \lim_{T \rightarrow \infty} \frac{1}{T} E \left[\sum_{t=0}^{T-1} R^\Phi(S(t)) \right] = \pi^\Phi R^\Phi, \quad (1)$$

其中 π^Φ 是在策略 Φ 下系统的稳态概率. 次用户的优化目标是寻找最优的探测策略 Φ^* , 使相应的平均传输率最大, 即 $\eta^{\Phi^*} = \max_{\Phi} \eta^\Phi$.

在马尔可夫决策过程模型中, 泊松方程是分析和

求解的基础. 对于这个问题, 泊松方程可写作

$$\eta^\Phi + g^\Phi(s) = R^\Phi(s) + P^\Phi g^\Phi(s), \quad s \in \mathcal{S}, \quad (2)$$

其中 g^Φ 为满足泊松方程的解, 又被称为性能势, 它反映了当前探测在未来对信道状态及数据传输量的影响. 泊松方程的解并不唯一, 不同解之间相差一个常数项, 满足 $\pi^\Phi g^\Phi = 0$ 的特解为

$$g^\Phi(s) = \lim_{T \rightarrow \infty} E \left\{ \sum_{t=0}^{T-1} (R^\Phi(S(t)) - \eta^\Phi) | S(0) = s \right\}. \quad (3)$$

根据泊松方程 (2), 可以推导出不同策略下的性能 (平均传输率) 差分公式为

$$\eta^{\Phi'} - \eta^\Phi = \pi^{\Phi'} (R^{\Phi'} - R^\Phi) + \pi^{\Phi'} (P^{\Phi'} - P^\Phi) g^\Phi. \quad (4)$$

基于此, 可以得到机会式频谱接入优化问题的最优性条件及最优性方程^[7,9].

定理 1 (最优性条件及最优性方程) 策略 Φ^* 是最优策略, 当且仅当

$$R^{\Phi^*} + P^{\Phi^*} g^{\Phi^*} \geq R^\Phi + P^\Phi g^{\Phi^*}, \quad \forall \Phi, \quad (5)$$

即策略 Φ^* 满足

$$R^{\Phi^*} + P^{\Phi^*} g^{\Phi^*} = \max_{\Phi} \{R^\Phi + P^\Phi g^{\Phi^*}\}. \quad (6)$$

从最优方程可以看出, 最优策略既考虑当前时隙的数据传输量 R^Φ , 也考虑到当前探测对未来数据传输量的影响 $P^\Phi g^{\Phi^*}$. 根据最优性方程, 本文可以设计出寻找最优策略的迭代算法. 算法的核心思想是, 从任意初始策略 Φ_0 出发, 在第 l 步迭代时, 基于策略 Φ_l , 按照下式改进策略:

$$\Phi_{l+1} \in \arg \max \{R^\Phi + P^\Phi g^{\Phi_l}\}. \quad (7)$$

当 Φ_{l+1} 与 Φ_l 无差别时, 算法停止在最优策略.

以上便是基于标准马尔可夫决策过程模型的求解, 但模型的状态空间 $\mathcal{S} = \mathbf{N}^N$ 是可数无限的, 在实际应用中, 由于存储空间有限, 不可能保存所有状态信息. 另一方面, 当某个信道很久未被探测时, 即 $k \rightarrow \infty$ 时, 对次用户而言, 可认为信道的状态分布为极限分布, 即信道状态的分布 (λ_I, λ_B) 满足

$$(\lambda_I, \lambda_B) = (\lambda_I, \lambda_B) \begin{bmatrix} p_{II} & p_{IB} \\ p_{BI} & p_{BB} \end{bmatrix}.$$

因此, 此时不必关心很久之前所探测到的信道状态, 也就没有必要存储所有状态的信息. 综合这两点原因, 本文将引入事件的概念, 用来集结多个状态转移, 并在此基础上建立该问题基于事件的决策过程模型.

2 基于事件的优化方法

2.1 基于事件的决策过程模型

在基于事件的模型中, 事件被定义为状态转移的集合^[8]. 具体来说, 令 $\langle s, s' \rangle$ 表示系统从状态 s 到状态 s' 的转移, $s, s' \in \mathcal{S}$, 则所有状态转移的集合记为 $\mathcal{E} =$

$\{\langle s, s' \rangle : s, s' \in \mathcal{S}\}$; 而一个事件 e 就是 \mathcal{E} 的一个子集, 即 $e \subseteq \mathcal{E}$. 对于机会式频谱接入优化问题, 状态转移可以表示为 $\langle s, s' \rangle = \langle (k_1, k_2, \dots, k_N), (k'_1, k'_2, \dots, k'_N) \rangle$, 其中 k_i 和 k'_i 分别表示次用户在当前时隙探测前后对信道 i 的认识, $i = 1, 2, \dots, N$. 显然, 如果在当前时隙信道 i 被探测为空闲, 则 $k'_i = 1$; 如果信道 i 被探测为忙碌, 则 $k'_i = 2$; 而对于其他没有被探测的信道, $k'_j = k_j + 2, j = 1, 2, \dots, i-1, i+1, \dots, N$. 对于任意状态 $s = (k_1, k_2, \dots, k_N)$, 用 $A_i^I(s)$ 表示次用户探测到信道 i 为空闲后的系统状态, 即

$$A_i^I(s) =$$

$$(k_1 + 2, k_2 + 2, \dots, k_{i-1} + 2, 1, k_{i+1} + 2, \dots, k_N + 2);$$

用 $A_i^B(s)$ 表示次用户探测到信道 i 为忙碌后的系统状态, 即

$$A_i^B(s) =$$

$$(k_1 + 2, k_2 + 2, \dots, k_{i-1} + 2, 2, k_{i+1} + 2, \dots, k_N + 2).$$

令 $A_i(s) = \{A_i^I(s), A_i^B(s)\}$ 表示在状态 s 探测信道 i 后系统可能出现的状态集合. 为了叙述方便, 引入常数 K , 对于任意状态 $s = (k_1, k_2, \dots, k_N)$, 若 $k_i \leq 2K$, 则认为信道 i 最近被探测过; 否则, 认为信道 i 长时间未被探测过.

在机会式频谱接入问题中, 事件分为可观事件、可控事件与自然转移事件. 顾名思义, 可观事件是可以被观测到的状态转移的集合. 当系统状态为 $s = (k_1, k_2, \dots, k_N)$ 时, 此时的可观事件为当前状态到所有可能状态的转移的集合, 记为

$$O(s) = \{\langle s, s' \rangle : s' \in A_i(s), i = 1, 2, \dots, N\}. \quad (8)$$

由于实际存储和计算条件的限制, 本文限定可观事件数量为有限个. 如果所有信道在最近都被探测过, 即 $k_i \leq 2K$, 则将可观事件记为

$$O(k_1, k_2, \dots, k_N) = O_e, \quad e = (e_1, e_2, \dots, e_N),$$

$$e_i = k_i \leq 2K, \quad i = 1, 2, \dots, N.$$

如果有多个信道 j_1, j_2, \dots, j_n 长时间未被探测过, 即 $k_{j_m} > 2K, m = 1, 2, \dots, n$, 则将可观事件记为

$$O(k_1, k_2, \dots, k_N) = O_e, \quad e = (e_1, e_2, \dots, e_N),$$

$$e_{j_m} = 2K + 1, \quad m = 1, 2, \dots, n,$$

$$e_i = k_i \leq 2K, \quad i \neq j_m, \quad i = 1, 2, \dots, N.$$

不难看出, 可观事件的集合为 $\mathcal{E}^O = \{O_e : e = (e_1, e_2, \dots, e_N), e_i = 1, 2, \dots, 2K + 1, i = 1, 2, \dots, N\}$. 由此, 原问题中可数无限个状态所对应的状态转移, 已经被划分成有限 $((2K + 1)^N)$ 个可观事件.

相对于可观事件, 可控事件是可控制其发生概率的事件. 在该问题中, 次用户的探测行为引起的状态转移为可控事件, 对信道 i 进行探测的可控事件记为

$$C_i = \{(s, s') : s \in \mathcal{S}, s' \in A_i(s)\}. \quad (9)$$

因此, 可控事件的集合为 $\mathcal{E}^C = \{C_i : i = 1, 2, \dots, N\}$, 即可控事件 C_i 与探测信道 i 一一对应.

除了可观事件与可控事件, 第3类事件被称为自然转移事件, 这类事件中的状态转移由系统固有属性决定且发生的概率不受控制. 在这个问题中, 在系统状态为 $s = (k_1, k_2, \dots, k_N)$, $k_j > 2K$, 次用户选定某个信道 i 进行探测后, 系统会转移到 $A_i(s)$ 的其中一个状态 ($A_i^I(s)$ 或 $A_i^B(s)$). 具体哪个状态由系统本身属性决定, 本质是由所探测的信道的状态(空闲或者忙碌)决定的. 当次用户观测到可观事件 O_e , 并选择可控事件 C_i 发生后, 如果所探测信道的状态为空闲, 则自然转移事件记为

$$T(I|O_e, C_i) = \langle s, A_i^I(s) \rangle; \quad (10)$$

如果所探测信道的状态为忙碌, 则自然转移事件记为

$$T(B|O_e, C_i) = \langle s, A_i^B(s) \rangle. \quad (11)$$

通过引入事件的概念, 传统马尔可夫决策过程模型中的每个状态转移都可以看作是某个可观事件、某个可控事件及某个自然转移事件3者的交集. 例如, 在某个时隙, 信道 j 长时间未被探测过, 即当前状态为 $s = (k_1, k_2, \dots, k_N)$, $k_j > 2K$, 那么, 此时可以观测到可观事件 $O(s) = O_e$, $e = (k_1, k_2, \dots, k_{j-1}, 2K+1, k_{j+1}, \dots, k_N)$ 发生, 系统的具体状态未知, 即只知道 $k_j > 2K$, 但不知道 k_j 的具体值. 随后次用户选择信道 i 进行探测, 即选择可控事件 C_i 发生. 然后, 系统根据固有属性(所探测信道的状态), 发生自然转移事件 $T(I|O_e, C_i)$ 或 $T(B|O_e, C_i)$, 转移到新的状态 s' . 不难看出, 如果所探测信道的状态为空闲, 则

$$\langle s, s' \rangle = \langle s, A_i^I(s) \rangle = O(s) \cap C_i \cap T(I|O(s), C_i);$$

如果所探测信道的状态为忙碌, 则

$$\langle s, s' \rangle = \langle s, A_i^B(s) \rangle = O(s) \cap C_i \cap T(B|O(s), C_i).$$

在基于事件的决策过程模型中, 在第 t 个时隙, 当某个可观事件 $O(S(t)) \in \mathcal{E}^O$ 发生之后, 次用户依照基于事件的策略 Ψ 选择信道 i 进行探测, 即决定可控事件 C_i 的发生概率 $P^\Psi(C_i|O(S(t)))$, $i = 1, 2, \dots, N$. 这里基于事件的策略 Ψ 是从可观事件集合 \mathcal{E}^O 到信道集合 $\{1, 2, \dots, N\}$ 的映射. 在策略 Ψ 下, 次用户在该时隙数据传输量的期望值为 $R^\Psi(O(S(t)))$, 同时根据系统固有属性(取决于所探测信道的状态, 与策略无关), 或依概率 $P(I|O(S(t)), C_i)$ 发生自然转移事件 $T(I|O(S(t)), C_i)$, 转移到新的状态 $A_i^I(S(t))$; 或依概率 $P(B|O(S(t)), C_i)$ 发生自然转移事件 $T(B|O(S(t)), C_i)$, 转移到新的状态 $A_i^B(S(t))$. 之后系统进入第 $t+1$ 个时隙, 重复上述过程. 在基于事件策略 Ψ 下, 次用户的长期平均数据传输率记为 η^Ψ , 次用户的优化目

标是寻找最优的基于事件的探测策略 Ψ^* , 使得相应的平均传输率最大, 即 $\eta^{\Psi^*} = \max_{\Psi} \eta^\Psi$.

2.2 基于事件的性能差分公式

下面推导基于事件的决策过程模型中的性能(平均传输率)差分公式. 设 $\pi^\Psi(O_e)$ 表示在策略 Ψ 下, 系统平稳时可观事件 O_e 发生的概率; $\pi^\Psi(s|O_e)$ 表示在策略 Ψ 下系统平稳时, 可观事件 O_e 发生的条件下, 系统状态为 s 的条件概率. 由全概率公式可知

$$\pi^\Psi(s) = \sum_{O_e \in \mathcal{E}^O} [\pi^{\Psi'}(s|O_e)\pi^{\Psi'}(O_e)]. \quad (12)$$

定义

$$R^{\Psi'}(O_e) = \sum_{s \in \mathcal{S}} \pi^{\Psi'}(s|O_e)R^{\Psi'}(s), \quad (13)$$

$$R^{\Psi', \Psi}(O_e) = \sum_{s \in \mathcal{S}} \pi^{\Psi'}(s|O_e)R^\Psi(s), \quad (14)$$

$$g^\Psi(A_i(s)) = P(I|O(s), C_i)g^\Psi(A_i^I(s)) + P(B|O(s), C_i)g^\Psi(A_i^B(s)), \quad (15)$$

$$g^\Psi(O_e, C_i) = \sum_{s \in \mathcal{S}} \pi^\Psi(s|O_e)g^\Psi(A_i(s)), \quad (16)$$

$$g^{\Psi', \Psi}(O_e, C_i) = \sum_{s \in \mathcal{S}} \pi^{\Psi'}(s|O_e)g^\Psi(A_i(s)). \quad (17)$$

考虑两个不同的基于事件的策略 Ψ' 和 Ψ , 两个策略下的平均传输率满足如下定理.

定理 2 基于事件策略的性能差分公式

$$\begin{aligned} \eta^{\Psi'} - \eta^\Psi = & \sum_{O_e \in \mathcal{E}^O} \pi^{\Psi'}(O_e)[R^{\Psi'}(O_e) - R^{\Psi', \Psi}(O_e)] + \\ & \sum_{O_e \in \mathcal{E}^O} \pi^{\Psi'}(O_e) \left\{ \sum_{C_i \in \mathcal{E}^C} [P^{\Psi'}(C_i|O_e) - P^\Psi(C_i|O_e)] \times \right. \\ & \left. g^{\Psi', \Psi}(O_e, C_i) \right\}. \end{aligned} \quad (18)$$

证明 根据基于状态的标准马尔可夫决策过程模型中的差分公式(4), 有

$$\begin{aligned} \eta^{\Psi'} - \eta^\Psi = & \sum_{s \in \mathcal{S}} \pi^{\Psi'}(s)(R^{\Psi'}(s) - R^\Psi(s)) + \\ & \sum_{s \in \mathcal{S}} \pi^{\Psi'}(s) \sum_{s' \in \mathcal{S}} [P^{\Psi'}(s'|s) - P^\Psi(s'|s)]g^\Psi(s'). \end{aligned} \quad (19)$$

下面分别考虑式(19)等号右边的两部分. 根据全概率公式或式(12), 对于第1部分, 有

$$\begin{aligned} & \sum_{s \in \mathcal{S}} \pi^{\Psi'}(s)(R^{\Psi'}(s) - R^\Psi(s)) = \\ & \sum_{s \in \mathcal{S}} \left\{ \sum_{O_e \in \mathcal{E}^O} [\pi^{\Psi'}(s|O_e)\pi^{\Psi'}(O_e)](R^{\Psi'}(s) - R^\Psi(s)) \right\} = \\ & \sum_{O_e \in \mathcal{E}^O} \pi^{\Psi'}(O_e) \left[\sum_{s \in \mathcal{S}} \pi^{\Psi'}(s|O_e)(R^{\Psi'}(s) - R^\Psi(s)) \right] = \\ & \sum_{O_e \in \mathcal{E}^O} \pi^{\Psi'}(O_e)[R^{\Psi'}(O_e) - R^{\Psi', \Psi}(O_e)]. \end{aligned}$$

类似地, 对于第 2 部分, 有

$$\begin{aligned} & \sum_{s \in \mathcal{S}} \pi^{\Psi'}(s) \sum_{s' \in \mathcal{S}} [P^{\Psi'}(s'|s) - P^{\Psi}(s'|s)] g^{\Psi}(s') = \\ & \sum_{s \in \mathcal{S}} \left\{ \sum_{O_e \in \mathcal{E}^O} [\pi^{\Psi'}(s|O_e) \pi^{\Psi'}(O_e)] \times \right. \\ & \left. \sum_{C_i \in \mathcal{E}^C} [(P^{\Psi'}(C_i|O_e) - P^{\Psi}(C_i|O_e))] \times \right. \\ & \left. [P(I|O_e, C_i) g^{\Psi}(A_i^I(s)) + P(B|O_e, C_i) g^{\Psi}(A_i^B(s))] \right\} = \\ & \sum_{O_e \in \mathcal{E}^O} \pi^{\Psi'}(O_e) \left\{ \sum_{C_i \in \mathcal{E}^C} (P^{\Psi'}(C_i|O_e) - P^{\Psi}(C_i|O_e)) \times \right. \\ & \left. \left[\sum_{s \in \mathcal{S}} \pi^{\Psi'}(s|O_e) g^{\Psi}(A_i(s)) \right] \right\}. \end{aligned}$$

将两部分合并, 即可得式 (18). \square

定理 2 比较了不同策略的数据传输率, 是进行性能评价和优化的重要基础.

3 基于样本路径的策略迭代算法

下面从实际应用的角度, 结合机会式频谱接入优化问题的特点, 讨论基于样本路径的基于事件策略的迭代算法.

3.1 近似处理

尽管定理 2 或式 (18) 可以给出两个策略平均数据传输率的差异, 但由于式中的 $R^{\Psi', \Psi}$ 和 $g^{\Psi', \Psi}$ 依赖于策略 Ψ' 和 Ψ , 在进行策略迭代时, 无法只根据当前策略 Ψ 找到更好的策略 Ψ' . 另一方面, 由于存储空间有限, 不可能保存所有状态的单位时隙数据传输量 $R^{\Psi}(s)$ 和性能势 $g^{\Psi}(s)$, 因此需要考虑近似的求解方法. 根据可观事件的定义, 如果所有信道在最近都被探测过, 则根据此时的可观事件可以确定当前的系统状态, 即此时的可观事件 O_e 对应于唯一的系统状态 $s_0 = (k_1, k_2, \dots, k_N)$, 满足 $k_i = e_i \leq 2K$, $i = 1, 2, \dots, N$, 且与所采用的策略无关, 即有

$$\pi^{\Psi'}(s_0|O_e) = \pi^{\Psi}(s_0|O_e) = 1, O(s_0) = O_e.$$

此时, 式 (18) 中的 $R^{\Psi', \Psi}(O_e)$ 可以转化为

$$\begin{aligned} R^{\Psi', \Psi}(O_e) &= \sum_{s \in \mathcal{S}} \pi^{\Psi'}(s|O_e) R^{\Psi}(s) = \\ \pi^{\Psi'}(s_0|O_e) R^{\Psi}(s_0) &= \pi^{\Psi}(s_0|O_e) R^{\Psi}(s_0) = \\ \sum_{s \in \mathcal{S}} \pi^{\Psi}(s|O_e) R^{\Psi}(s) &= R^{\Psi}(O_e). \end{aligned}$$

类似地, 有

$$\begin{aligned} g^{\Psi', \Psi}(O_e, C_i) &= \sum_{s \in \mathcal{S}} \pi^{\Psi'}(s|O_e) g^{\Psi}(A_i(s)) = \\ \pi^{\Psi'}(s_0|O_e) g^{\Psi}(A_i(s_0)) &= \pi^{\Psi}(s_0|O_e) g^{\Psi}(A_i(s_0)) = \\ \sum_{s \in \mathcal{S}} \pi^{\Psi}(s|O_e) g^{\Psi}(A_i(s)) &= g^{\Psi}(O_e, C_i). \end{aligned}$$

如果当前时隙有一个或多个信道长时间未被探测过, 则根据此时的可观事件 O_e 无法确定当前的系

统状态 s , 即系统可能处于多个不同的状态; 但观测到的可观事件均为 O_e , 记系统此时可能的状态集合为 $V(O_e) = \{s : O(s) = O_e\}$. 根据可观事件的定义可知, $V(O_e)$ 与所采用的策略无关, 即有

$$\sum_{s \in V(O_e)} \pi^{\Psi'}(s|O_e) = \sum_{s \in V(O_e)} \pi^{\Psi}(s|O_e) = 1.$$

在实际中, 这种有信道长期未被探测使用的情况的发生概率较小, 因此本文可对这种情况作近似处理. 文献 [7] 中已经证明, 在有信道长期未被探测的情况下, 系统所有可能的状态之间的距离很近. 根据函数的连续性, 假设在这种情况下, 系统所有可能状态的单位时隙数据传输量期望值和性能势均近似相等. 设 \bar{s} 为集合 $V(O_e)$ 中的任一状态, 则式 (18) 中

$$\begin{aligned} R^{\Psi', \Psi}(O_e) &\approx R^{\Psi}(\bar{s}) \sum_{s \in V(O_e)} \pi^{\Psi'}(s|O_e) = \\ R^{\Psi}(\bar{s}) \sum_{s \in V(O_e)} \pi^{\Psi}(s|O_e) &\approx R^{\Psi}(O_e); \end{aligned} \quad (20)$$

类似地

$$\begin{aligned} g^{\Psi', \Psi}(O_e, C_i) &\approx g^{\Psi}(A_i(\bar{s})) \sum_{s \in V(O_e)} \pi^{\Psi'}(s|O_e) = \\ g^{\Psi}(A_i(\bar{s})) \sum_{s \in V(O_e)} \pi^{\Psi}(s|O_e) &\approx g^{\Psi}(O_e, C_i). \end{aligned} \quad (21)$$

经过近似处理之后, $R^{\Psi}(O_e)$ 看作在策略 Ψ 下基于事件的单位时隙数据传输量, 而 $g^{\Psi}(O_e, C_i)$ 看作在策略 Ψ 下基于事件的集结性能势. 基于事件的性能差分公式 (18) 可以近似为

$$\begin{aligned} \eta^{\Psi'} - \eta^{\Psi} &\approx \\ \sum_{O_e \in \mathcal{E}^O} \pi^{\Psi'}(O_e) [R^{\Psi'}(O_e) - R^{\Psi}(O_e)] &+ \\ \sum_{O_e \in \mathcal{E}^O} \pi^{\Psi'}(O_e) \left\{ \sum_{C_i \in \mathcal{E}^C} (P^{\Psi'}(C_i|O_e) - P^{\Psi}(C_i|O_e)) \times \right. \\ \left. g^{\Psi}(O_e, C_i) \right\}. \end{aligned} \quad (22)$$

这种近似所带来误差的上界为

$$\begin{aligned} \Delta &\leq \\ \sum_{O_e \in \mathcal{E}_K^O} \pi^{\Psi'}(O_e) \left(\max_{s \in V(O_e)} R^{\Psi}(s) - \min_{s \in V(O_e)} R^{\Psi}(s) \right) &+ \\ \sum_{O_e \in \mathcal{E}_K^O} \pi^{\Psi'}(O_e) \left(\max_{s \in \mathcal{S}} g^{\Psi}(s) - \min_{s \in \mathcal{S}} g^{\Psi}(s) \right). \end{aligned} \quad (23)$$

其中: $\mathcal{E}_K^O = \{O_e : \exists e_i = 2K + 1, i = 1, 2, \dots, N\}$, 即为有多个信道长时间未被探测过的可观事件的集合. 显然, 随着 K 的增大, $\sum_{O_e \in \mathcal{E}_K^O} \pi^{\Psi'}(O_e)$ 会不断减小, 因

此这种近似带来的误差也会不断缩小. 当所考虑的策略为有限个时, 如果误差小于任意两个策略的平均传输率之差, 误差可以忽略, 策略迭代算法仍能收敛到最优策略.

假设所讨论的任意基于事件的策略 Ψ 对应的稳态概率分布 π^Ψ 均满足遍历性条件 $\pi^\Psi(O_e) > 0, \forall O_e \in \mathcal{E}^O$. 根据近似差分公式 (22), 可以设计出基于事件策略的迭代算法. 其基本思想是, 从任意初始的基于事件的策略 Ψ_0 出发, 在第 l 步迭代时, 基于策略 Ψ_l , 按照下式改进策略:

$$\Psi^{l+1} \in \arg \max_{\Psi} \{R^\Psi + P^\Psi g^{\Psi^l}\}. \quad (24)$$

当 Ψ_{l+1} 与 Ψ_l 的性能之差满足误差要求时, 算法停止.

注 1 上述对于基于事件策略的遍历性假设: $\pi^\Psi(O_e) > 0, \forall O_e \in \mathcal{E}^O$, 并不苛刻. 对于不满足上述假设的策略 Ψ , 可以按照类似在线学习算法(如强化学习)中常用的 ϵ 贪婪策略构造其邻近策略 Ψ_ϵ ^[10]: 取微小正数 $\epsilon > 0$, 在每个时隙, 首先依照均匀分布产生一个 0 到 1 之间的随机数 r , 如果 $r > \epsilon$, 则策略 Ψ_ϵ 选择探测的信道与 Ψ 相同; 否则, 策略 Ψ_ϵ 在所有信道中等概率随机选择. 这样构造的策略 Ψ_ϵ 可满足前面遍历性的假设, 而其平均传输率又与 Ψ 相差无几, 满足 $|\eta^{\Psi_\epsilon} - \eta^\Psi| \leq \epsilon \eta^{\Psi^*}$, 其中 η^{Ψ^*} 为最优策略传输率. 下面将讨论上述策略迭代算法基于样本路径的实现, 这也要求系统从长期来看总能以一定概率访问所有的事件与行动的组合, 即满足遍历性假设; 否则, 没有数据就无法进行统计学习.

3.2 基于样本路径的迭代算法

上述基于灵敏度分析的方法的另一个优点是系统与样本路径结合紧密. 不难发现, 基于事件的数据传输量和性能势都可以从系统运行时的样本路径上估计得到. 由此可以设计出基于样本路径的策略迭代算法. 这个算法对比较复杂的系统, 或系统的参数未知的情况, 都非常实用.

首先, 考虑基于事件的单位时隙数据传输量的期望值 $R^\Psi(O_e)$ 的估计. 设系统所发生的可观事件的序列为 $\{E(1), E(2), \dots\}$. 根据近似式 (20), 可得如下估计式:

$$R^\Psi(O_e) = \frac{\sum_{t=0}^{T-1} I[O_e|E(t)]R^\Psi(E(t))}{\sum_{t=0}^{T-1} I[O_e|E(t)]}, \quad (25)$$

其中指示函数 $I[O_e|E(t)] = 1$, 当且仅当 $E(t) = O_e$. 类似地, 根据集结性能势的近似式 (21), 有

$$g^\Psi(O_e, C_i) = \frac{\sum_{t=0}^{T-M+1} I[O_e|E(t)]I[C_i|\Psi(E(t))] \sum_{m=0}^{M-1} R^\Psi(E(t+m))}{\sum_{t=0}^{T-M+1} I[O_e|E(t)]I[C_i|\Psi(E(t))]}, \quad (26)$$

其中 $I[C_i|\Psi(E(t))] = 1$, 当且仅当 $\Psi(E(t)) = C_i$.

基于这些估计式, 可以设计出如下迭代算法.

算法 1 基于样本路径的基于事件策略迭代算法.

Step 1: 初始化. 选择合适的算法参数, 包括 K, T, M 等, 并选取任意策略 Ψ_0 作为初始策略, 令 $l = 0$.

Step 2: 策略评价. 在策略 Ψ_l 下, 观测系统运行 T 步, 根据式 (25) 和 (26) 估计相关参量.

Step 3: 策略改进. 根据式 (24) 构造一个更好的策略, 记为 Ψ_{l+1} .

Step 4: 如果改进后策略的性能满足 $|\eta^{\Psi_{l+1}} - \eta^{\Psi_l}| < \epsilon$, 其中 ϵ 为误差要求, 则算法终止; 否则, 令 $l = l + 1$, 并返回 Step 2.

下面用 1 个仿真示例来验证算法的效果.

例 1 考虑包含两个信道的通信系统, 信道的状态转移概率为 $P_{BI} = 0.8, P_{II} = 0.3$. 这种情况下, 贪婪策略被证明是最优策略^[5], 即只根据 $P_{BI} > P_{II}$ 决定所探测的信道. 若当前时隙被探测的信道的状态为忙碌 (B), 则下个时隙仍然探测该信道; 若当前时隙探测被探测的信道的状态为空闲 (I), 则在传输数据之后, 下个时隙探测另一条信道. 在下面的仿真中, 本文将上述最优策略与基于样本路径的迭代算法进行比较.

在基于样本路径的迭代算法中, 假设并不知道信道的状态转移概率. 选择 $K = 5, T = 10^5$. 图 1 中, 实线是基于样本路径的策略迭代算法经过 10 次迭代的结果, 虚线是最优策略的平均传输率. 由图 1 可见, 迭代策略的平均传输率很快收敛, 且逼近最优传输率, 这说明基于样本路径的策略迭代算法是有效的.

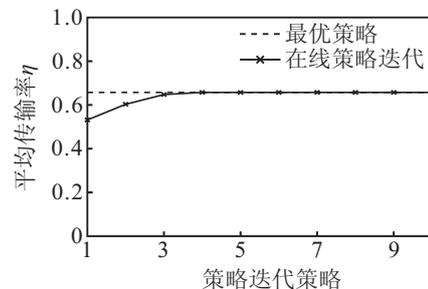


图 1 基于样本路径的策略迭代结果与最优策略

进一步, 本文根据样本路径统计出有信道长期未被探测使用的所有事件的概率约为 0.007, 这与有信道长期未被探测使用的事件为小概率事件的假设相符. 由式 (23) 计算得到基于事件的近似处理带来的误差的上界为 $\Delta \leq 0.007(1 + 4.2) \approx 0.036$. 与迭代策略的平均传输率 0.66 相比, 误差的影响小于 $0.036/0.66 \approx 5.5\%$, 这说明这种近似处理是合理的.

4 结 论

本文考虑机会式频谱接入技术中次用户对可利用信道进行探测与接入策略的优化问题. 通过引入事件的定义, 将含有无限个状态的原问题被转化为包含有限个事件的决策问题. 基于灵敏度分析的方法, 推导出基于事件策略的性能差分公式; 通过合理的近似, 得到基于事件策略的迭代算法, 并给出了近似处理的误差上限, 算法的误差可以根据实际需要进行控制. 最后, 通过仿真示例验证了所提出算法的有效性与近似处理的合理性. 进一步的研究工作可以将基于事件的优化方法推广到更一般的模型, 如考虑多个次用户, 对信道探测有误差, 以及信道之间相关关联或状态转移概率不同等情况.

参考文献(References)

- [1] Zhao Q, Sadler B M. A survey of dynamic spectrum access[J]. *IEEE Signal Processing Magazine*, 2007, 24(3): 79-89.
 - [2] Zhao Q, Tong L, Swami A, et al. Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework[C]. *Proc of 1st IEEE Int Symposium on New Frontiers in Dynamic Spectrum Access Networks*. Baltimore: IEEE Press, 2005: 589-600.
 - [3] Zhao Q, Krishnamachari B, Liu K Q. Low-complexity approaches to spectrum opportunity tracking[C]. *Proc of 2nd Int Conf Cognitive Radio Oriented Wireless Networks and Communications*. Orlando: IEEE Press, 2007: 27-35.
 - [4] Zhao Q, Krishnamachari B, Liu K Q. On myopic sensing for multi-channel opportunistic access: Structure, optimality and performance[J]. *IEEE Trans on Wireless Communications*, 2008, 7(12): 5431-5440.
 - [5] Ahmad S H, Liu M, Javidi T, et al. Optimality of myopic sensing in multi-channel opportunistic access[J]. *IEEE Trans on Information Theory*, 2009, 55(9): 4040-4050.
 - [6] Zhao Q, Geirhofer S, Tong L, et al. Opportunistic spectrum access via periodic channel sensing[J]. *IEEE Trans on Signal Processing*, 2008, 56(2): 785-796.
 - [7] 黄永皓, 陈曦. 机会式频谱接入优化问题的研究[J]. *控制与决策*, 2010, 25(6): 857-861.
(Huang Y H, Chen X. A study on opportunistic spectrum access optimization problem[J]. *Control and Decision*, 2010, 25(6): 857-861.)
 - [8] Cao X R. Basic ideas for event-based optimization of Markov systems[J]. *Discrete Event Dynamic Systems*, 2005, 15(2): 169-197.
 - [9] Cao X R. Stochastic learning and optimization: A sensitivity-based approach[M]. New York: Springer, 2007: 387-446.
 - [10] Sutton R S, Barto A G. Reinforcement learning: An introduction[M]. Cambridge: The MIT Press, 1998: 122.
-
- (上接第1642页)
- [3] 杨文, 侍洪波, 汪小帆. 卡尔曼一致滤波算法综述[J]. *控制与决策*, 2011, 26(4): 481-488.
(Yang W, Shi H B, Wang X F. A survey of consensus based Kalman filtering algorithm[J]. *Control and Decision*, 2011, 26(4): 481-488.)
 - [4] Spanos D P, Saber R O, Murray M R. Approximate distributed Kalman filtering in sensor networks with quantifiable performance[C]. *Proc of the 4th Int Symposium on Information Processing in Sensor Networks*. Los Angeles: IEEE Press, 2005: 133-139.
 - [5] Saber R O. Distributed Kalman filtering algorithms for sensor networks[C]. *Proc of the 46th IEEE Conf on Decision and Control*. New Orleans: IEEE Press, 2007: 5492-5498.
 - [6] Saber R O, Jalalkamali P. Coupled distributed estimation and control for mobile sensor networks[J]. *IEEE Trans on Automatic Control*, 2012, 57(10): 2609-2614.
 - [7] Stanković S S, Stanković M S, Stipanović D M. Consensus based overlapping decentralized estimation with missing observations and communications faults[J]. *Automatica*, 2009, 45(6): 1397-1406.
 - [8] Yu W W, Chen G R, Wang Z D, et al. Distributed consensus filtering in sensor networks[J]. *IEEE Trans on Systems, Man and Cybernetics, Part B: Cybernetics*, 2009, 39(6): 1568-1577.
 - [9] Su H S, Chen G R, Wang X F, et al. Adaptive second-order consensus of networked mobile agents with nonlinear dynamics[J]. *Automatica*, 2011, 47(2): 368-375.
 - [10] Zavlanos M M, Tanner H G. Hybrid control for connectivity preserving flocking[J]. *IEEE Trans on Automatic Control*, 2009, 54(12): 2869-2875.