

基于演化博弈的用户接入机制

王月平^{1*}, 徐涛²

(1. 南京航空航天大学 计算机科学与技术学院, 南京 211106; 2. 中国民航大学 计算机科学与技术学院, 天津 300300)

(* 通信作者电子邮箱 84545773@qq.com)

摘要: 用户接入问题就是无线终端选择接入某个服务基站的问题。用户接入可以看作无线资源管理的第一步, 对于网络性能有着重要的影响, 在实现负载均衡、控制干扰、提高频谱和能量效率等方面起着非常重要的作用。针对包含宏基站和全双工小基站的多层异构网络特点, 考虑了分离多接入机制, 即允许一个终端在上行和下行接入到不同的多个基站, 从而实现性能的提升。在此基础上, 将异构网络中的上行下行分离多接入的用户接入问题建模成一个演化博弈问题, 其中用户是博弈方, 相互之间进行资源的竞争, 基站的接入选择就是博弈中的策略, 每个用户希望通过策略的选择实现自身效用的最大化。此外, 基于演化博弈和强化学习设计了低复杂度自组织用户接入算法, 用户可以根据当前的策略选择收益来进行策略调整, 并最终达到均衡状态, 实现了用户公平性。最后通过大量的仿真实验验证了所提方法的有效性。

关键词: 用户接入; 异构网络; 上行下行分离; 演化博弈

中图分类号: TP393; TN92 **文献标志码:** A

User association mechanism based on evolutionary game

WANG Yueping^{1*}, XU Tao²

(1. College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing Jiangsu 211106, China;

2. School of Computer Science and Technology, Civil Aviation University of China, Tianjin 300300, China)

Abstract: User association is the problem that a wireless terminal chooses to access one serving base station. User association can be seen as the first step in wireless resource management, which has an important impact on network performance, and plays a very important role in achieving load balance, interference control, improvement of spectrum and energy efficiency. Aiming at the characteristics of multi-layer heterogeneous network including macro base stations and full-duplex small base stations, a separate multiple access mechanism was considered, which means allowing a terminal access different and multiple base stations in the uplink and downlink, so as to realize the performance improvement. On this basis, the user association problem with separation of uplink and downlink multi-access in heterogeneous network was modeled into an evolutionary game problem, in which the users act as the players to perform the resource competition with each other, the access choices of base stations are strategies in the game, and every user wants to obtain the maximum of own effectiveness by the choice of strategy. Besides, a low-complex self-organized user association algorithm was designed based on evolutionary game and reinforcement learning. In the algorithm, the user was able to adjust the strategy according to the revenue of current strategy choice, and reached an equilibrium state in the end, realizing user fairness. Finally, extensive simulations were performed to verify the effectiveness of the proposed method.

Key words: user association; heterogeneous network; separation of uplink and downlink; evolutionary game

0 引言

随着网络结构的密集化、异构化,以及全双工等新技术的引入,许多无线通信网络中的基础问题需要进行重新研究,用户接入(user association 或 cell association)便是其中之一^[1-2]。在无线通信网络特别是超密集异构网络中,一个无线终端通常处于多个基站的服务范围内。用户接入问题就是为无线终端选择接入某个基站进行服务和数据传输的问题,无论采用何种无线技术以及网络部署方式,用户接入机制都是不可或缺的。同时用户接入机制作为无线资源管理的第一步,对于网络性能有着重要的影响,在实现负载均衡、干扰控制、提高频谱和能量效率等方面起着非常重要的作用^[3]。

异构网络中的用户接入机制可以分为网络端驱动的集中

式设计 Network-driven,即由网络中心节点作出接入决策;以及用户端驱动的分布式设计 User-driven,即由用户终端作出接入决策。网络端驱动的集中式设计的优点就是中心节点能够收集网络的整体信息,因此可以作出针对系统整体的最优接入决策,代价就是具有较高的复杂度。相比而言,用户端驱动的用户接入机制的优势则具有较低的复杂度。两种方案各有其优缺点,适用于不同的网络结构及节点类型。在超密集异构网络中,大部分的小基站由不同的组织、机构、甚至个人(如家庭基站)进行部署。针对这种情况,将面向包含宏基站和全双工小基站的多层异构自组织网络研究用户端驱动的低复杂度用户接入机制。

现有的用户接入机制的研究^[3-8]与设计几乎都是基于上行下行关联的,即用户终端在上行和下行必须接入到同一个

收稿日期: 2019-11-04; 修回日期: 2020-01-13; 录用日期: 2020-01-14。

作者简介: 王月平(1971—),男,北京人,博士研究生,主要研究方向:无线网络; 徐涛(1965—),男,天津人,教授,博士,CCF会员,主要研究方向:计算机网络。

基站^[9]。对于单层同构网络来说,这种接入机制简单而有效。首先,采用上下行关联的接入机制便于网络的设计与运行,特别是便于信令同步、资源管理等;其次,在单层同构网络中,能提供最好下行连接的基站通常也能提供最好的上行连接。而在多层异构网络中,存在着不同层次基站之间以及上行下行之间的不对称问题,即不同层次的基站之间以及上行下行之间通常在发射功率、覆盖范围、信道质量、数据流量、回程容量以及硬件等方面有较大差别,并且这种不对称会随着各层基站密度的增加而更加突出。这时基于上行下行关联的用户接入将会严重制约系统的性能,如负载均衡、干扰、频谱和能量效率等。例如,当采用基于最强下行信号且上下行关联的用户接入机制时,因为宏基站与小基站发射功率存在较大差异,该机制将会使大部分用户都连接到宏基站,从而造成严重的负载不均衡,并降低了小基站部署的意义。此外基于频谱效率的考虑,宏基站与小基站通常复用同一频谱资源,这时会存在以下情况:基于最强下行信号用户连接到较远的宏基站;然而在进行上行数据传输时,该用户需要使用较大的传输功率来达到其上行性能要求,但这样会对附近的小基站造成严重的上行干扰,从而影响网络的整体性能并增加自身的能量消耗。为此,上行下行分离的概念被提出^[10],即上行传输和下行传输可以选择接入不同的基站。上行下行关联的用户接入可以看作是上下行分离用户接入的一种特例,因此,从理论上说上下行分离可以带来更好的性能。初步的仿真研究^[11-13]已经证明在两层异构网络中,简单的上行下行分离机制可以带来较大的性能提升,例如,可以提高上行传输速率、降低发射功率和上行干扰,以及实现更好的负载均衡等。目前只有少量研究面向超密集异构网络。考虑到超密集网络中基站密度大、用户接入选择多的特点,一些学者提出了采用双连接^[14]或者多连接^[15]的用户接入机制,即用户可以同时连接到两个或者多个基站。上述工作目前只研究了最简单的多接入机制,即选择接入最近的 k 个基站,但是没有作进一步的优化设计。

本文考虑允许一个用户在上行和下行接入多个不同的基站,将自组织分离多接入问题建模成演化博弈问题,并用一组微分方程(即复制者动态)来描述用户接入策略的选择调整过程;提出了演化均衡作为博弈的均衡解,并分析了其稳定性;此外,基于强化学习和深度强化学习设计了自组织用户接入算法,用户可以根据当前的策略选择收益来进行策略调整,并最终达到均衡状态,实现了用户之间的公平性。

1 系统建模

1.1 网络模型

考虑由1个宏基站、 K 个小基站和 N 个用户组成的双层异构网络,其中小基站可以进行全双工传输。假设网络中的小基站和用户在空间上都是随机分布,在所考虑的网络中,将使用提出的分离多接入机制,其中用户只在下行时可以选择接入宏基站。此外,假设用户在上行时只能选择接入1个小基站,而下行时可以选择多个小基站。两个节点 i 和 j 之间的信道增益 $G_{i,j}$ 可以定义为:

$$G_{i,j} = |h_{i,j}|^2 d_{i,j}^{-\xi} \quad (1)$$

其中: $|h_{i,j}|^2$ 表示节点 i 与 j 之间的由于衰减和阴影效应等造成的复合信道功率增益, $d_{i,j}$ 表示 i 到 j 之间的距离, ξ 是路径损失指数。接下来定义一个二进制变量 X_{nk} 来表示小基站 k 服务用户 n 的情况,此外,用 S_{nk}^b 来表示信道接入概率,假设基站给连接的用户以同等的概率分配信道资源。在这种情况下,信道接入概率 S_{nk}^b 可被表示为:

$$S_{nk}^b = \begin{cases} 1 / \sum_{n'=1}^N X_{n'k}, & \sum_{n'=1}^N X_{n'k} > 0 \\ 0, & \text{其他} \end{cases} \quad (3)$$

式(3)表示只有接入概率大于0,用户才能成功访问该信道进行数据传输。

1.2 干扰模型

下面分析一下系统中的干扰情况。与传统的上行下行耦合的用户接入方法相比,分离多接入不可避免会导致节点之间产生额外的干扰,这又会影响节点最后的性能。本节分别分析上行和下行传输中的干扰,具体如下:

上行链路传输中的干扰 上行链路传输中,用户作为发送端,小基站作为接收端,必须分析小基站收到的干扰情况。每一个小基站在上行和下行允许许多用户同时接入,同时,对于某个用户来说,能在上行和下行连接到多个不同的小基站。在这样的系统中,在小基站接收到的与上行干扰包括3个部分:来自其他进行下行传输全双工小基站的干扰、来自其他进行上行传输的用户的干扰,以及自干扰,具体表示为:

$$I_k^{\text{UL}} = I_{k'k}^{\text{DL}} + I_{n'k}^{\text{UL}} + I_k^{\text{self}} \quad (4)$$

其中: $I_{k'k}^{\text{DL}}$ 表示由其他进行下行传输的小基站 k' 引起的干扰, $I_{n'k}^{\text{UL}}$ 表示由其他进行上行传输的用户 n' 引起的干扰, I_k^{self} 表示小基站 k 的自干扰,具体来说:

$$I_{k'k}^{\text{DL}} = \sum_{k' \in K \setminus \{k\}} P_{k'} G_{k'k} \sum_{n'=1}^N x_{n'k'}^{\text{DL}} S_{n'k'}^b \quad (5)$$

$$I_{n'k}^{\text{UL}} = \sum_{n' \in N \setminus \{n\}} P_{n'} G_{n'k} (1 - x_{n'k}^{\text{DL}}) \sum_{k'=1}^K x_{n'k'}^{\text{UL}} S_{n'k'}^b \quad (6)$$

$$I_k^{\text{self}} = P_k / \delta \quad (7)$$

下行链路传输中的干扰 用户作为下行链路传输的接收者会受到其他节点传输的干扰。类似的,对连接到小基站 k 的UE(User Equipment)的干扰进行分析,干扰也包含3个部分,具体表示如下:

$$I_n^{\text{DL}} = I_{k'n}^{\text{DL}} + I_{n'n}^{\text{UL}} + I_n^{\text{self}} \quad (8)$$

其中:

$$I_{k'n}^{\text{DL}} = \sum_{k' \in K \setminus \{k\}} P_{k'} G_{k'n} (1 - x_{n'k'}^{\text{DL}}) \sum_{n'=1}^N x_{n'k'}^{\text{DL}} S_{n'k'}^b \quad (9)$$

$$I_{n'n}^{\text{UL}} = \sum_{n' \in N \setminus \{n\}} P_{n'} G_{n'n} \sum_{k'=1}^K x_{n'k'}^{\text{UL}} S_{n'k'}^b \quad (10)$$

$$I_n^{\text{self}} = P_n / \delta \quad (11)$$

1.3 数据传输速率

在所考虑的网络中,将用户和小基站之间数据传输速率作为用户接入的性能衡量指标。显然,一个用户更希望接入到能够提供更高传输速率的小基站。影响传输速率的关键因素是小基站能提供的带宽、发送功率,以及节点之间的干扰。对于给定的小基站 k ,用户 n 连接到 k 的上行和下行传输速率可以用Shannon-Hartley定理计算:

$$\Phi_{nk} = B_{nk} \log(1 + \text{SINR}_{nk}) \quad (12)$$

其中 SINR_{nk} 是信号干扰噪声比。对于分离多接入,有:

$$\text{SINR}_{n,k}^{\text{UL}} = P_n \frac{G_{nk}}{I_{k'k}^{\text{DL}} + I_{n'k}^{\text{UL}} + I_k^{\text{self}} + \sigma^2} \quad (13)$$

$$\text{SINR}_{n,k}^{\text{DL}} = P_k \frac{G_{nk}}{I_{k'n}^{\text{DL}} + I_{n'n}^{\text{UL}} + I_n^{\text{self}} + \sigma^2} \quad (14)$$

其中 σ^2 是加性高斯白噪声(Additive White Gaussian Noise, AWGN),假设对所有用户都是一样的。这样,用户 n 在上行和下行连接到小基站 k 的上行和下行速率可以相应表达为:

$$\Phi_{nk}^{\text{DL}} = B_{nk} \text{lb} \left(1 + P_k \cdot \frac{G_{nk}}{I_{k'n}^{\text{DL}} + I_{n'n}^{\text{UL}} + I_n^{\text{self}} + \sigma^2} \right) \quad (15)$$

$$\Phi_{nk}^{\text{UL}} = B_{nk} \text{lb} \left(1 + P_n \cdot \frac{G_{nk}}{I_{k'k}^{\text{DL}} + I_{n'k}^{\text{UL}} + I_k^{\text{self}} + \sigma^2} \right) \quad (16)$$

2 演化博弈建模及分析

本文将基于演化博弈来设计用户端驱动的多接入机制,演化博弈论是研究有限理性博弈方之间相互作用的数学工具。在演化博弈论中,博弈方根据自己的适应度(即收益)选择策略。在无线网络中,博弈方可以是一定区域内争夺无线资源的 N 个用户,这也构成了群体。用 $n_i(t)$ 表示 t 时刻选择纯策略 $l \in L$ 的代理人数量,其中 L 为纯策略集合。标准化变量 $x_l(t) = n_l(t)/N$ 表示策略 l 的群体份额,此处

$x_l(t) \in [0, 1]$, $\sum_{l=1}^{|L|} x_l(t) = 1$ 。所有纯策略 $|L|$ 的群体份额构成种群状态,可以用向量 $\mathbf{x}_l(t) = [x_1(t), x_2(t), \dots, x_{|L|}(t)]^T \in X$ 表示,其中 X 是状态空间。 $\pi(l, x)$ 表示使用纯策略 l 的玩家的期望收益,也表示适应度。因此群体的平均收益可以推导为 $\pi(x, x) = \sum_{l=1}^{|L|} x_l \pi(l, x)$,动态策略适应过程(即演化博弈中的选择机制)可用以下微分方程描述的复制者动态方程来建模:

$$\dot{x}_l = \delta x_l (\pi(l, x) - \pi(x, x)) \quad (17)$$

其中 δ 是群体的学习速率,控制着选择变化的速度。根据复制者动态,预期收益高于平均收益的策略的总体份额将增加,而预期收益低于平均收益的策略的总体份额将减少。当所有的群体份额不变时,演化均衡就会达到(i.e., $\dot{x}_l = 0$ 对于所有的 $l \in L$)。演化均衡是沿着动态轨迹的不动点,也就是说,当群体状态演化到这样的固定点时,群体状态将保持不变。演化均衡的物理意义是,在这个均衡中,没有一个个体想要改变策略因为它的收益等于总体的平均收益。虽然可能存在多种演化均衡,但并不是所有的均衡都是稳定的。

本文采用演化博弈的动机如下:

有限理性假设 在传统非合作博弈论中通常假设博弈方都是完全理性的,即博弈方完全清楚在某个策略组合下自己的收益和别人的收益,并选择策略追求自身收益的最大化。这是一个非常强的假设,在很多实际情况无法满足该假设,而演化博弈只要求博弈方具有有限理性,即在了解部分收益的情况下,通过不断学习调整策略,逐渐提高自己的收益。

博弈动态描述 演化博弈可以对群体之间各博弈方之间的动态交互进行描述建模。一个博弈方可以观察到其他博弈方的行为,并从观察中学习,然后再作出策略调整。此外,通过复制者动态,任意时间 t 的博弈状态可以被确定,这样可以用来研究博弈方策略调整的一个轨迹或者趋势。

筛选博弈均衡解 在传统博弈论中,纳什均衡是最常用的解的形式。纳什均衡能保证在其他博弈方法不偏离均衡策略时,一个博弈方不能单独通过调整自己的策略来获得更高的收益。然而,在许多情况下,一个博弈可能有多个纳什均衡解,因此,就需要进行均衡解的筛选。在演化博弈中,演化均衡提供了一种筛选方法,并且能保证筛选过后的解的稳定性。

低复杂度学习机制 在演化博弈中,各博弈可以通过一些低复杂度的学习机制来达到演化均衡,这在许多对算法复杂度要求较高的实际问题中很有用处。

公平性 在演化博弈中,在通过不断动态调整达到演化均衡状态时,选择不同策略会得到同样收益,实现了公平性。

2.1 问题建模

考虑用户接入的策略动态调整过程,建模演化博弈问题 $\mathcal{G} = (N, S, \pi)$,具体介绍如下:

博弈方 在本文的系统模型中,每个能接入到小基站的活跃用户都是一个博弈方。用 $\mathcal{N} = \{1, 2, \dots, N\}$ 来表示本博弈中的博弈方的集合。

群体 在演化博弈中,博弈方的集合构成了一个群体。

策略集合 每个博弈方的策略就是选择要接入的小基站。需要注意的是,同一群体中的博弈方拥有相同的策略集。

群体比例 指在一个群体中,选择某个策略的博弈方数量占整个群体数量的比例,即 $x_i = n_i/N$,其中 $i \in S$, n_i 是选择策略 i 的博弈方数量。显然 $\sum_{i=1}^S x_i = 1$ 。

群体状态 一个群体中的所有策略的群体比例构成了群体状态,可以通过 $[x_1, x_2, \dots, x_i, \dots, x_s] \in X$ 来表示,其中 X 表示在给定网络模型下包含了所有可能群体状态的状态空间。

收益函数 每个用户作为博弈方需要根据收益来调整其接入策略。在本博弈中,定义一个博弈方的收益函数为接入速率和接入费用之间的加权差。其中在不同策略下的用户速率可以表示为 $\mathbf{r} = [r_1, r_2, \dots, r_i, \dots, r_s]$ 。根据系统模型部分的数据传输速率公式,可以得到:

$$r_i(t) = \sum_{k^n} B_{nk}^{\text{UL}} \text{lb} (1 + \text{SINR}_{n,k}^{\text{UL}}) + \sum_{k^d} B_{nk}^{\text{DL}} \text{lb} (1 + \text{SINR}_{n,k}^{\text{DL}}) \quad (18)$$

其中: k^n, k^d 表示用户 n 在上行和下行连接到的基站集合,相应的速率是连接到每个小基站的速率之和。基于此,一个博弈方选择某个策略的收益函数可以定义如下:

$$\pi_i(t) = \lambda r_i(t) - c(b_i, p_i) \quad (19)$$

其中: λ 代表收益权重, $c(b_i, p_i)$ 是用户接入小基站的费用函数,与分配的带宽 b 和发射功率 p 相关,具体定义如下:

$$c(b_i, p_i) = \mu b_i + \nu p_i \quad (20)$$

其中 μ 和 ν 分别是单位带宽和功率的费用。

2.2 接入策略动态选择

正如博弈建模阶段所述,用户接入是一个动态选择过程,每个博弈方在时间 t 会根据其收到的效益 $\pi_i(t)$ 以及当前整个群体的平均收益来动态调整其策略,这一过程不断重复,从而不断地提升自己的收益,直到达到均衡状态。可以用复制者动态来对这个策略动态调整的过程建立数学模型。复制者动态是一个常微分方程表达如下:

$$\dot{x}_i(t) = \gamma x_i(t) [\pi_i(t) - \bar{\pi}(t)] \quad (21)$$

其中: $\dot{x}_i(t)$ 是 $x_i(t)$ 关于 t 的微分,代表选择策略 i 的群体比例 x_i 的变化率, $\dot{x}_i(t) > 0$ 时表示选择该策略的博弈方将增加; γ 是设定的学习率,用来控制用户接入策略调整的频率; $\bar{\pi}(t)$ 是同一群体中所有博弈方的平均收益,可以根据式(22)计算得出:

$$\bar{\pi}(t) = \sum_{i=1}^N \pi_i(t) \cdot x_i(t) \cdot N^{-1} \quad (22)$$

从式(22)可以看出,一个策略的群体比例变化率与选择该策略得到的收益与整个群体平均收益成正比。

需要注意的是,在给定初始条件 $\sum_i x_i(0) = 1$,可以保证在任一时刻 $t \in [0, \infty]$ 都能满足如下约束:

$$\sum_i x_i(t) = 1; x_i(t) \in [0, 1] \quad (23)$$

即使使用复制者动态来描述群体比例的动态调整的过程中,所有策略的群体比例之和始终为1。为了验证这一点,令 $f_i(x) = \gamma x_i(t) [\pi_i(t) - \bar{\pi}(t)]$ 。首先可以得到 $\sum_i f_i(x) = 0$,即所

有群体比例之和不会随时间变化。此外,还可以得到当 $x_i(t) = 0$ 时 $f_i(x) \geq 0$, 当 $x_i(t) = 1$ 时 $f_i(x) \leq 0$, 所以可以保证 $x_i(t) \in [0, 1]$ 。

2.3 演化均衡、演化稳定策略分析

复制者动态描述了所有的群体比例变化率。从用户接入的复制者动态方程可以看出,当采用某个策略的收益比群体平均收益要高时(即 $\pi_i(t) > \bar{\pi}(t)$, $\pi_i(t) - \bar{\pi}(t) > 0$), 相应的 $\dot{x}_i(t) > 0$, 这表示采用策略 i 的用户数量会相应增加。考虑演化均衡作为该演化博弈的均衡解。演化博弈可以定义为上述复制者动态中的不动点, 即 $\dot{x}_i(t) = 0, \forall i$ 。根据定义, 演化均衡可以通过求解一组代数方程 $\dot{x}_i = 0, \forall i$ 得到。

当达到演化均衡时, 群体中选择不同策略能得到相等的收益(即 $\pi_i(t) = \bar{\pi}(t)$), 因此一方面保证了公平性, 另一方面, 在达到均衡时, 没有博弈方有动机来改变选择的策略。需要注意的是, 根据上面方法求解代数方程可能得到多组演化均衡, 其中包括边界均衡点(即 $\exists i, x_i = 1$)以及内点均衡点(即 $x_i \in [0, 1], \forall i$)。对于本文定义的用户接入演化博弈来说, 边界均衡解不是一个稳定的解, 任何小的扰动都会让系统偏离这个均衡解。相比而言, 可以证明所求的内点均衡是渐近稳定并且是一个演化稳定策略(Evolutionary Stable Strategy, ESS)^[8]解。下面首先介绍一下演化稳定策略解。

如果对于任一群体状态 $\in X$ 且 $y \neq x$, 存在 $\varepsilon_y \in (0, 1)$ 使得 $\varepsilon \in (0, \varepsilon_y)$ 不等式 $\pi(x, \varepsilon_y + (1 - \varepsilon)x) > \pi(y, \varepsilon_y + (1 - \varepsilon)x)$ 成立, 则均衡群体状态 $x^* \in X$ 是演化稳定的。演化稳定策略(ESS)对入侵者具有很强的抵抗能力, 即群体中所有的突变体最终都会消失。对于ESS群体分布, 状态的微小扰动并不影响收敛到均衡态。物理意义是, 扰动态的平均效益小于均衡态的平均效益。这样, ESS对入侵者(即扰动)具有很强的抵抗力。实际上, 在某些情况下, 复制动态是全局渐近稳定的。在这种情况下, 任何偏离均衡的地方都会收敛回均衡。

首先来证明建立的用户接入演化博弈是一个稳定群体博弈, 因为下面条件对所有群体状态 $x \neq y$ 都满足:

$$(x_i(t) - y_i(t))(\pi_i(x) - \pi_i(y)) \leq 0 \quad (24)$$

因此, 所建模的演化博弈具有如下均衡是中性稳定(neutrally stable)的属性。此外, 根据复制者动态的特征方程, 其系数矩阵总是有 $2M$ 个负特征值, 这样表明所建立的复制者动态具有渐近稳定属性(asymptotically stable property)。因此, 用户接入动态系统能够从任何非边界初始群体状态渐进演化到一个内点演化均衡点, 因此可以知道所得到的内点演化均衡点是一个演化稳定策略解(ESS)。

3 基于演化博弈的分离多接入算法设计

针对基于演化博弈的异构网络中分离多接入, 本文设计了两种具体的实现算法: 第一种方法是基于强化学习, 其中每个用户尝试不同的接入策略, 观察其得到收益, 并在必要的条件下尝试策略调整; 第二种是在强化学习的机基础上引入了深度学习, 基于深度强化学习改进了第二种算法。下面具体介绍这两种算法。

3.1 基于强化学习的方法

在无中心节点条件下, 每个用户需要独立地通过尝试、观察和学习来调整策略, 并最终达到均衡状态。针对这一情况, 基于Q学习(强化学习的一种, 如图1所示)设计了一个无需其他用户信息的分布式算法(见算法1)。在算法中, 使用一个Q值来维护每个策略的相关信息。

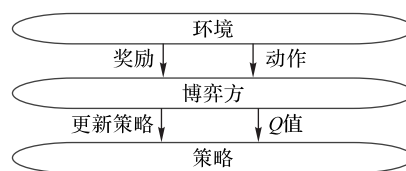


图1 强化学习架构

Fig. 1 Framework of reinforcement learning

具体算法描述如下:

算法1 基于强化学习的用户接入算法设计。

Input 初始Q值。

Output 选择策略。

- 1) 初始化Q值, 设置 $Q_i(0) = 0$, 设置循环变量 $iteration = 1$
- 2) Loop
- 3) if $rand() < \gamma$ & $iteration \leq Max-loop$ then
- 4) 随机选择一个策略 i //强化学习的探索阶段
- 5) else
- 6) 选择接入策略 $i^* = \arg \max_i Q_i(k)$ //强化学习的探索阶段
- 7) end if
- 8) 用户计算其收益 π_i
- 9) 更新Q值:

$$Q_i(k+1) = (1 - \lambda)Q_i(k) + \lambda(\pi_i + \beta \max_i Q_i(k))$$
- 10) end Loop

在基于强化学习的用户接入算法中, 一个用户以概率 γ 执行探索阶段, 相应的以概率 $1 - \gamma$ 执行利用阶段。 λ 是学习速率, 用来控制Q值的调整速率。Q值代表用户对未来迭代过程中的预期收益, Q值的更新依赖于之前的Q值与新观察到的收益值。在这一更新迭代过程中, 用户通过与环境交互学习逐渐提高其收益。

3.2 基于深度强化学习的方法

传统的强化学习局限于动作空间和样本空间都很小。当网络规模增加, 可选基站数量也将会大幅度增加, 从而一个用户的可选策略将会极大增加。上述强化学习方法难以有效解决。因此, 针对大状态空间和策略空间的情况, 进一步考虑了基于深度强化学习的方法。深度强化学习将深度学习的感知能力和强化学习的决策能力相结合, 深度强化学习原理框架如图2所示。

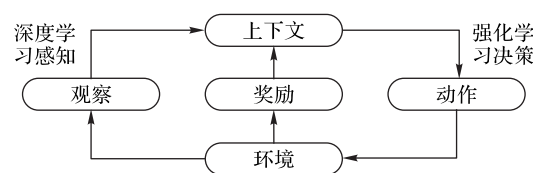


图2 深度强化学习架构

Fig. 2 Architecture of deep reinforcement learning

其学习过程可以描述为:

- 1) 在每个时刻 Agent 与环境交互得到一个高维度的观察, 并利用深度学习方法来感知观察, 以得到具体的状态特征表示;
- 2) 基于预期回报来评价各动作的价值函数, 并通过某种策略将当前状态映射为相应的动作;
- 3) 环境对此动作做出反应, 并得到下一个观察, 通过不断循环以上过程, 最终可以得到实现目标的最优策略。

基于深度强化学习的算法和上述强化学习的一大区别就是采用一个深度网络代表价值函数, 代替了强化学习中的Q

表。然后依据强化学习中的奖励值,为深度网络提供目标值,对网络不断更新直至收敛。本文使用的人工神经网络(Artificial Neural Network, ANN)架构如图3所示。具体算法过程与基于强化学习算法类似,因此不再赘述。

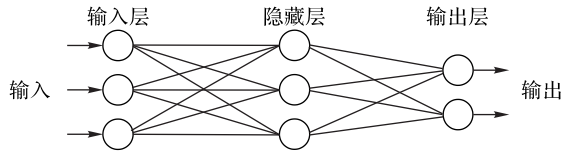


图3 ANN架构
Fig. 3 Framework of ANN

4 实验与结果分析

为了验证所提方法,本文在Matlab中进行仿真。具体考虑一个包含1个宏基站、2个小基站与10个用户的异构网络,基站与用户都随机分布在一个半径100 m的区域内,同时,考虑每个基站有不同的信道资源。

图4展示了复制者动态的收敛性,可以观察到,经过迭代后,不同策略群体比例最终也从初始状态达到了一个稳定的均衡状态。在达到均衡点后,策略选择的变化率为0,也就是说选择其他策略对于该用户并不能提高收益,因此没有偏离均衡策略的动机。这个均衡点也是复制者动态的不动点。此外,能够观察到初始占比较高的策略,在达到均衡时占比不一定很高,它反映出了提出的动态接入算法的公平性。

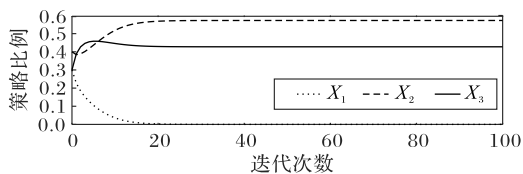


图4 群体比例演进

Fig. 4 Population proportion evolution

在图5中,以用户选择策略 l 的收益为例,验证了所提出的两种算法(基于强化学习的算法以及基于深度强化学习的算法)的收敛性,并与传统基于群体状态演进算法进行了比较。可以看到:经过迭代三种算法都收敛到了一个均衡状态,并获得了相应的收益;基于深度强化学习的方法能较快地收敛,并且能够获得较高的收益,而基于群体状态演进和基于强化学习的算法最终能获得相同的收益。

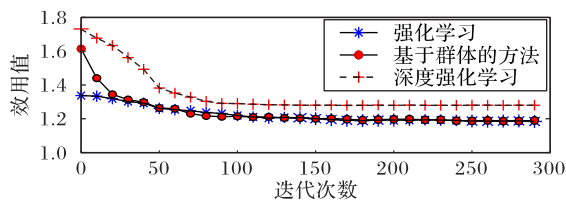


图5 不同算法比较(策略 $i=1$)

Fig. 5 Comparison of different algorithms ($i=1$)

5 结语

在本文中,研究了自组织网络的用户接入优化问题。针对多层异构网络特点,提出了分离多接入机制。本文将自组织分离多接入问题建模成演化博弈问题,其中用户可以根据当前的策略选择收益来进行策略调整,并最终达到均衡状态。使用复制者动态对这一策略动态调整进行了数学建模。基于群体状态演进、强化学习、深度强化学习设计了自组织用户接入算法。通过仿真验证了复制者动态以及所提出算法的收敛

性并比较了三种算法的性能。

参考文献 (References)

- [1] LIU D, WANG L, CHEN Y, et al. User association in 5G networks: a survey and an outlook[J]. IEEE Communications Surveys and Tutorials, 2017, 18(2):1018-1044.
- [2] ANDREWS J G, BUZZI S, CHOI W, et al. What will 5G be?[J]. IEEE Journal on Selected Areas in Communications, 2014, 32(6): 1065-1082.
- [3] ZHANG T, ZHU Q. Game-based TDMA MAC protocol for vehicular network[J]. Journal of Communications and Networks, 2017, 19(3):209-217.
- [4] ZHANG Y, YANG C, LI J, et al. Distributed interference-aware traffic offloading and power control in ultra-dense networks: mean field game with dominating player[J]. IEEE Transactions on Vehicular Technology, 2019, 68(9): 8814-8826.
- [5] SEO J B, KWON T, CHOI J. Evolutionary game approach to uplink NOMA random access systems[J]. IEEE Communications Letters, 2019, 23(5): 930-933.
- [6] JIE Y, KAMAL A E, ALNUEN M A. Cooperative multi-path routing solution with real-time optimization for streaming application using auction theory[J]. IEEE Access, 2019, 7: 112520-112534.
- [7] 王俊社, 蒋彤彤, 张琛, 等. 密集小蜂窝网络中基于用户接入的能效优化研究[J]. 计算机工程, 2019, 45(12): 45-51, 70. (WANG J S, JIANG T T, ZHANG C, et al. Research on energy efficiency optimization based on user access in dense small cell network[J]. Computer Engineering, 2019, 45(12):45-51, 70.)
- [8] 徐大专, 邓昭. 多用户接入与SLT编码的二维喷泉系统及性能分析[J]. 数据采集与处理, 2018, 33(6): 945-952. (XU D Z, DENG Z. Two-dimensional multiuser access and SLT coding fountain system and its performance[J]. Journal of Data Acquisition and Processing, 2018, 33(6):945-952.)
- [9] HOSSAIN E, HASSAN M. 5G cellular: key enabling technologies and research challenges[J]. IEEE Instrumentation and Measurement Magazine, 2015, 18(3): 11-21.
- [10] WANG X, HAIDER F, GAO X, et al. Cellular architecture and key technologies for 5G wireless communication networks[J]. IEEE Communications Magazine, 2014, 52(2):122-130.
- [11] BOCCARDI F, HEATH R W, LOZANO A, et al. Five disruptive technology directions for 5G[J]. IEEE Communications Magazine, 2015, 52(2):74-80.
- [12] ELSHAER H, BOCCARDI, DOHLERH F. Downlink and uplink decoupling: a disruptive approach design for 5G networks[C]// Proceedings of the 2014 IEEE Global Communications Conference. Piscataway: IEEE, 2014: 1798-1803.
- [13] ZHENG K, YANG Z, ZHANG K, et al. Big data-driven optimization for mobile networks toward 5G[J]. IEEE Network, 2016, 30(1):44-51.
- [14] AHMAD S A, DATLA D. Distributed power allocation in heterogeneous networks with dual connectivity using backhaul state information[J]. IEEE Transactions on Wireless Communications, 2015, 14(8):4574-4581.
- [15] KAMEL M I, HAMOUDA W, YOUSSEF A M. Multiple association in ultra-dense networks[C]// Proceedings of the 2016 IEEE International Conference on Communications. Piscataway: IEEE, 2016: 1-6.

WANG Yueping, born in 1971, Ph. D. candidate. His research interests include wireless networking.

XU Tao, born in 1965, Ph. D., professor. His research interests include computer network.