doi:10.11959/j.issn.1000-436x.2015240

星地融合网络中基于 Q 学习的切换算法研究

熊丹妮，李屹
（北京邮电大学 泛网无线通信教育部重点实验室，北京 100876）

摘 要：基于地面辅助基站（ATC）的星地融合网络（MSS-ATC）具有覆盖范围广、用户体验佳的特点，切换机制是该融合网络主要研究的问题之一。针对卫星链路时延大、卫星网用户速度范围广的特点，综合考虑了用户接收信号强度（RSS）和用户运动速度，提出了一种基于卡尔曼滤波和 Q 学习的切换决策算法。比较了所提算法与传统算法在链路衰减率、切换次数和网络收益的性能，实验结果表明所提算法在性能上得到了很大的提升，并且能很好地适应高速运动状态。

关键词：切换；MSS-ATC；RSS 预测；Q 学习
中图分类号：TN927 **文献标识码**：A

Q-learning based handoff algorithm for satellite system with ancillary terrestrial component

XIONG Dan-ni, LI Yi
(Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract*:* In the integrated satellite-terrestrial communication system, i.e., mobile satellite system with ancillary terrestrial component (MSS-ATC), the long transmission delay of satellite link was a huge challenge which may lead to high handoff dropping probability. In order to address this problem, a novel handoff decision strategy was proposed based on the predictive RSS and Q-learning algorithm. Extensive simulation results demonstrate that the proposed scheme can decrease the handoff dropping probability, reduce the unnecessary handoff times and maximize the network reward. In addition, the proposed scheme can also adapt to the situation of high-speed movement very well.
Key words: handoff scheme; MSS-ATC; predictive RSS; Q-learning

1 引言

移动卫星通信系统具有顽健性强、覆盖范围广的特点，广泛应用于应急通信服务[1]。由于卫星通信链路依赖于视距传输，为减少阴影物阻挡的影响，星地融合网络成为近年来研究的热点[2]。地面辅助基站（ATC, ancillary terrestrial component）作为通信系统辅助，被广泛应用于星地融合网络中，不仅有效地改善了高楼林立的城区的通信质量，也更大限度地利用了卫星网络的频谱资源[3,4]。

基于 ATC 的移动卫星系统（MSS-ATC, mobile satellite system- ancillary terrestrial component）包括卫星网络和地面网络 2 部分，卫星网的中继为 GEO 卫星，地面网的基站为 ATC 基站。卫星网用户按运行速度可以分为低速终端和高速终端，所谓低速终端，是指移动速度在每小时几十公里以下，如手持终端、车载终端等；所谓高速终端，是指一些特殊行业中时速达成百上千的终端，如航天飞机、宇宙飞船、火箭导弹等；而实际上，目前地面用户的速率分布范围已经十分大，从几公里每小时的行人到几百公里每小时的高铁。由于 MSS-ATC 系统具有卫星链路时延长、用户运动速度范围大的特点，网络切换面临极大的挑战，特别是速度在几十到几百公里每小时的地面中低速用户。近年来，切换问题受到众多学者的关注，但多数研究都是针对于地面异构网络和 LEO 卫星系统，因而现有切换机制还不能很好地满足 MSS-ATC 系统的需求。

传统切换决策算法一般是基于接收信号强度

收稿日期：2014-06-30；修回日期：2015-05-12

（RSS, received signal strength）的最优化问题。然而，若只考虑接收信号强度，当用户处于小区边缘时，容易产生乒乓效应[5]。为解决这类问题，文献[6]提出了一种基于 MDP 的自适应切换算法，采用 RSS 预测和磁滞门限的机制有效地改善了乒乓效应，提高了系统性能，但是该方法没有考虑到用户速度范围广的问题。文献[7,8]提出了基于 RSS 和用户速度的切换算法，但是这些研究都是在低速率的场景下，不能保证高速率场景下的切换成功率和减少切换次数。文献[9~13]对基于多属性决策的切换算法进行了研究，但这些研究都是基于当前的用户状态决策，没有考虑到下一时刻用户状态的变化。以上所有方法对切换问题都做出了深入研究并提高了系统性能，但是由于 MMS-ATC 系统独有的时延长、用户速度范围大的特点，现有方法都不能很好地应用于其中，因此需要专门针对 MSS-ATC 的特点做出优化。Q 学习算法[14]是由 Walkins 在 1989 年提出的一种强化学习算法，与动态规划算法相似，通过不断与环境交互，时刻以最大化长期收益为目标，学习最佳决策集，决策结果不仅依赖于决策时刻状态，也依赖于接下去的状态。

本文综合考虑了上述的 MSS-ATC 系统面临的挑战，针对速率范围在几十到几百公里每小时的中低速地面用户，提出了一种基于 RSS 预测和 Q 学习算法的切换决策算法。该算法由 2 部分组成，首先，采用卡尔曼滤波算法预测用户下一时刻的接收信号强度；然后，根据预测信息，利用 Q 学习算法做出切换决策。采用 RSS 预测，能使切换提前进行，改善了由于卫星链路长时延带来的切换问题；采用 Q 学习算法，将用户当前决策与长期收益联系起来，使切换决策更加准确。除此之外，该算法还考虑到用户速度，在高速率场景下，系统同样拥有很好的性能。

2　系统建模

本文研究 MSS-ATC 系统下的切换问题。MSS-ATC 系统由 GEO 卫星网络和 ATC 组成，GEO 卫星网络采用了多波束的技术，每个波束相当于一个宏蜂窝，波束间采用频谱七色复用技术以提高频谱利用率，一个波束内有多个 ATC 小区重叠覆盖某些区域，ATC 小区复用相邻波束的频带，使频谱资源再次复用。用户既可以跟卫星网通信也可以跟 ATC 基站通信[3]。一个波束内的典型系统架构如图 1 所示。

假定用户在移动，周期性检查邻近小区的 RSS。传统的切换机制是网络根据用户的测量报告做出切换决策。然而，简单地根据当前 RSS 做出的决策具有几个弊端：1）卫星链路的长时延会引起极高的切换掉线率；2）高速运动的用户在短时间内经过多个 ATC 小区，使其面临频繁的切换，增大系统开销；3）当用户运动至小区边缘，会受到乒乓效应的影响。

为解决这些问题，本文提出一种基于 RSS 预测和 Q 学习算法的切换决策算法。以磁滞门限为切换阈值，基于系统的预测值，采用 Q 学习方法提前做出切换决策，减小时延和乒乓效应的影响。同时对高速运动用户单独考虑，解决高速率场景下切换次数过多的问题。具体实现方法将在下文中详细讨论。

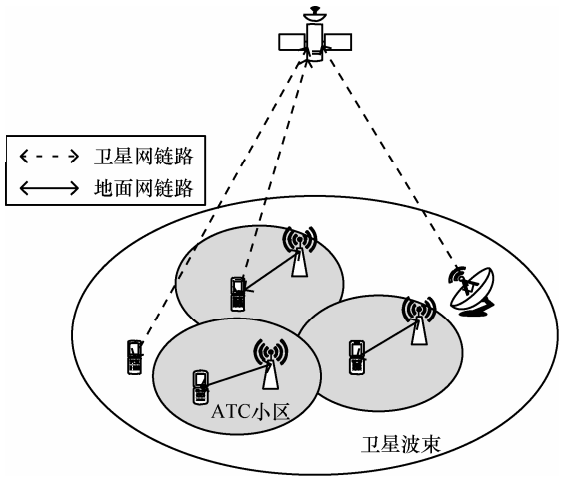


图 1　系统架构

3　基于 Q 学习的切换决策算法

算法模型如图 2 所示，切换决策过程分 2 步进行。首先，用户利用传感器测量自身速度、位置和邻近服务网络的 RSS，并将测量数据发送至当前已连接网络，网络根据测量信息以及用户的历史运动信息预测下一时刻的位置和速度，基于预测值计算下一时刻的 RSS；其次，获取预测信息和用户的历史状态信息后，网络端采用 Q 学习算法做出切换决策，并将最优决策发送给用户。

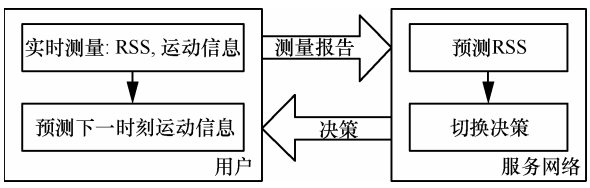


图 2　切换决策流程

3.1 RSS 预测介绍

在现有研究中，大部分切换模型都是基于 RSS 预测和磁滞门限[6,15]，采用这类方法可以有效地改善乒乓效应的影响，但是用户切换时刻的 RSS 可能已经减弱，不稳定的链路质量将导致切换掉线率升高。本文提出 RSS 预测的方法，首先利用卡尔曼滤波算法预测用户下一时刻的位置和速度，然后根据预测值计算 RSS，提前进行切换决策和执行切换动作。

3.1.1 卡尔曼滤波

卡尔曼滤波算法利用已有运动信息序列预测下一时刻的用户状态并及时修正模型参数以保证连续稳定的预测，该方法被广泛用于节点位置跟踪。二维空间内用户的运动方程为[16]

$$\boldsymbol{m}_{t+1}=\boldsymbol{\Phi}\boldsymbol{m}_t+w_t, w_t\sim \mathrm{N}(0,\boldsymbol{Q}) \tag{1}$$

$$\boldsymbol{z}_{t+1}=\boldsymbol{H}\boldsymbol{m}_t+\upsilon_t, \upsilon_t\sim \mathrm{N}(0,\boldsymbol{R}) \tag{2}$$

其中，$\boldsymbol{m}_t([x_t\ y_t\ v_{x,t}\ v_{y,t}]^{\mathrm{T}})$ 是用户运动状态，$\boldsymbol{\Phi}$ 为当前时刻 t 到下一时刻 t+1 的状态转移矩阵，$\boldsymbol{H}$ 是联系测量值和状态值的测量矩阵，随机数 w_t 和 υ_t 分别为与过程噪声和测量噪声有关的参数，它们相互独立且协方差矩阵分别为 $\boldsymbol{Q}$ 和 $\boldsymbol{R}$。

卡尔曼滤波跟踪的过程分为 2 步：用户预测和状态更新。其中用户预测过程分为 2 部分，用户运动状态预测和误差协方差预测[16]

$$\boldsymbol{m}_{t+1}^{-}=\boldsymbol{\Phi}\boldsymbol{m}_t \tag{3}$$

$$\boldsymbol{\Sigma}_{t+1}^{-}=\boldsymbol{\Phi}\boldsymbol{\Sigma}_t\boldsymbol{\Phi}^{\mathrm{T}}+\boldsymbol{Q} \tag{4}$$

在位置信息更新过程中，首先通过测量过程得到 $\boldsymbol{z}_{t+1}$，利用式（5）计算卡尔曼增益；然后综合预测状态和测量结果通过式（6）得到一个最佳的预测状态[17]

$$K_{t+1}=\frac{\boldsymbol{\Sigma}_{t+1}^{-}\boldsymbol{H}^{\mathrm{T}}}{\boldsymbol{H}\boldsymbol{\Sigma}_{t+1}^{-}\boldsymbol{H}^{\mathrm{T}}+\boldsymbol{R}} \tag{5}$$

$$\boldsymbol{m}_{t+1}=\boldsymbol{m}_{t+1}^{-}+K_{t+1}(\boldsymbol{z}_{t+1}-\boldsymbol{H}\boldsymbol{m}_{t+1}^{-}) \tag{6}$$

最后，通过式(7)得到后验误差协方差矩阵[17]

$$\boldsymbol{\Sigma}_{t+1}=(I-KI)\boldsymbol{\Sigma}_{t+1}^{-} \tag{7}$$

本设计对卡尔曼滤波的准确性进行了验证，假设一个节点在一个矩形区域内活动，图 3 显示了其实际位置和预测位置。仿真结果表明，卡尔曼滤波可以准确地跟踪到节点的运动情况，从而卡尔曼滤波结果可以用来预测用户下一时刻的 RSS 并帮助切换决策。

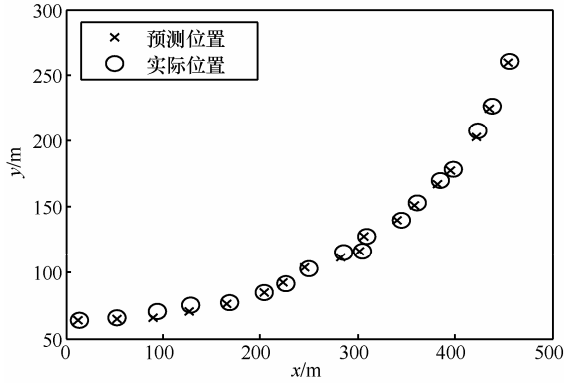


图 3 卡尔曼滤波预测位置与实际位置对比

3.1.2 RSS 预测

MSS-ATC 系统中，用户既可以跟 GEO 卫星通信也可以跟邻近 ATC 基站通信，用户将卡尔曼滤波的结果发送到当前服务网络，服务网络端通过这些节点信息预测下一时刻的 RSS。

用户 RSS 由 3 部分组成：路径损耗、阴影衰落和小尺度衰落。本文假设小尺度衰落可以通过接收端的滤波处理消除。对于 MSS 链路和 ATC 链路，采用一个通用路径损耗模型，在采样时刻 t 用户的 RSS 为

$$X_t=P_t+Z_t(\mathrm{dB}) \tag{8}$$

其中，P_t 为平均功率，其值由路径损耗决定，与发送信号强度和发送节点与接收节点间的距离相关。Z_t 为阴影衰落，大小主要由传播路径上的阻挡物（建筑物、树木等）决定。对于 RSS 的组成部分，ATC 链路和 MSS 链路有不同之处。

ATC 链路中的离散时间路径损耗模型为[18]

$$X_{\mathrm{ATC}}=\mu_{\mathrm{ATC}}-10\eta\log d_t+Z_{\mathrm{ATC}} \tag{9}$$

其中，μ_{ATC} 是平均接收信号强度，η 为传播常数，d_t 为用户到基站的距离，对于 ATC 系统，可以假设 Z_{ATC} 为一个常数。通过上文得到的下一时刻用户的位置 (x_t',y_t')，根据 ATC 的位置 $(x_{\mathrm{ATC}}^i,y_{\mathrm{ATC}}^i)$，可以估算下一时刻用户与 ATC 基站的距离[19]

$$d_t'=\sqrt{(x_t'-x_{\mathrm{ATC}}^i)^2+(y_t'-y_{\mathrm{ATC}}^i)^2} \tag{10}$$

预测的用户接收信号强度可以表示为

$$X_{\mathrm{ATC}}'=Z_{\mathrm{ATC}}+\mu_{\mathrm{ATC}}-10\eta\log\sqrt{(x_t'-x_{\mathrm{ATC}}^i)^2+(y_t'-y_{\mathrm{ATC}}^i)^2} \tag{11}$$

根据卫星信道的特点，衰落过程可以在理想信道状态和非理想信道状态之间进行切换，这种切换可以用一阶 Markov 两状态信道模型来描述[20]，每

个状态的统计概率密度函数服从 Loo 分布[21]。Markov 状态机根据转移概率矩阵，生成 Markov 状态序列，每个状态转移仅仅考虑当前状态和转移矩阵。设 Z_{GEO} 为基于该新信道模型的接收功率，代表了信道受阴影衰落影响的情况。由于用户到 GEO 卫星的距离基本不变，可以假设用户平均接收功率 μ_{GEO} 为常量。可以得到卫星链路的信道模型[19]

$$X_{\mathrm{GEO}} = \mu_{\mathrm{GEO}} + Z_{\mathrm{GEO}} \tag{12}$$

根据预测的位置和信道状态，可以预测 ATC 网络和卫星网络的 RSS，并依此做出切换决策。

3.2 基于 Q 学习的切换决策方案

在切换决策阶段，本设计视网络为一个学习者，通过不断地与环境变量交互，时刻以最大化长期收益为目标，使用 Q 学习算法[22]得到一个最佳决策集，决策结果不仅依赖于环境当前的状态，也依赖于接下去的状态以及相关的动作。

在 Q 学习方法中，学习者需要学习如何通过历史决策信息优化当前决策集。在决策时隙 t，学习者获取当前状态 s，然后执行动作 a，执行动作后，环境给学习者一个正向或负向反馈来表明动作是否正确，然后进入下一状态 s'。Q 学习的目的是通过多次决策过程为每个状态找到对应的最佳动作，形成最佳决策集 $\pi^*(s)\in A$ [14]，并最终使系统的总期望折扣回报最大。

3.2.1 系统状态和动作集

本文假设系统为有限状态离散时间动态系统，即假设用户所处的状态数是有限的，且在每个采样时刻，用户都可以选择接入不同网络。用 S 表示状态集，状态包括当前服务网络和邻近网络的 RSS，对应的网络编号 id 和用户相对于对应基站的速度 v。为了实现状态有限，将 RSS 和速度按大小划分并用具体数值表示大小等级，其中 $rss_i \in \{1,2,3,4,5\}$，$v_i \in \{-4,-3,-2,-1,1,2,3,4\}$，速度的负值表示用户正在朝着远离基站的方向运动。状态可以被表示为 $s=\{\{id_1,rss_1,v_1\},\cdots,\{id_4,rss_4,v_4\}\}$。动作由 a 表示，其取值为网络编号 $\{1,2,\cdots,N\}$，代表了切换决策，动作空间可以表示为 $A=\{a\mid a\in\{1,2,\cdots,N\}\}$。基于当前状态 $s_t\in S$，学习代理选择并执行一个动作 $a_t\in A$。执行动作后，用户以状态转移概率 $P_{s,s'}(a_t)$ 进入下一状态，并且获得一个值为 r_t 的回报，学习代理通过重复这样的学习过程会得到一个最佳决策集 $\pi^*(s)\in A$。

基于决策集 $\pi(s)\in A$，系统的总期望折扣回报为[22]

$$V^{\pi}(s)=\mathrm{E}\left[\sum_{t=0}^{\infty}\gamma_t r(s_t,x(s_t))\mid s_0=s\right] \tag{13}$$

其中，$\gamma_t\in[0,1)$ 为折扣因子，E 为期望运算符，$r(s_t,x(s_t))$ 为状态动作对的回报函数。可将式（13）化为贝叶曼等式的形式，基于最优策略 $\pi^*(s)\in A$，将问题化为求解贝叶曼的最佳准则问题[22]

$$V^*(s)=\max_{a\in A}\left[r(s,a)+\gamma\sum_{s'}P_{s,s'}(a)V^*(s')\right] \tag{14}$$

其中，折扣因子 $\gamma_t\in[0,1)$，$P_{s,s'}(a)$ 为在状态 s 执行动作 a，转换到 s' 的状态转移概率。在已知 $r(s,a)$ 和 $P_{s,s'}(a)$ 的情况下，可以求解式（14），然而 $P_{s,s'}(a)$ 的大小与无线网络环境参量有关，不易获取。使用 Q 学习算法时不需要知道 $r(s,a)$ 和 $P_{s,s'}(a)$ 的具体分布便能得到最优解，基于式（14），定义关于状态—动作对 (s,a) 的 Q 值函数[22]

$$M_q^{\pi}(s,a)=r(s,a)+\gamma\sum_{s'}P_{s,s'}(a)V^{\pi}(s') \tag{15}$$

在最优策略的作用下，定义[22]

$$M_q^{*}(s,a)=r(s,a)+\gamma\sum_{s'}P_{s,s'}(a)V^{*}(s') \tag{16}$$

其中，$V^*(s')=\max_{a\in A}[M_q^*(s,a)]$，算法根据每次执行动作和环境反馈，更新 Q 值函数。可以推算出

$$M_{q_{t+1}}(s,a)=M_{q_t}(s,a)+\rho[r(s,a)+\gamma\max_{a'\in A}(M_{q_t}(s',a'))-M_{q_t}(s,a)] \tag{17}$$

其中，学习率 $\rho=\dfrac{1}{1+T(s,a)}$，$T(s,a)$ 代表某个状态—动作对被访问的次数。显然，当 $T(s,a)$ 趋向于无穷时，ρ 趋向于 0，此时 $M_{q_t}(s,a)$ 收敛至 $M_q^*(s,a)$。通过反复的学习和决策过程，学习者得到最优决策集，得到最大的 Q 值。

3.2.2 Q 学习算法的收益函数

合理高效的回报函数使切换准确、快速，因而合理的回报函数是切换的必要条件。回报大小跟候选网络的 RSS、用户速度和切换动作这 3 个属性有关。不妨定义回报函数为[23]

$$r(s,a)=\omega_{rss}U_{rss}(rss_a)+\omega_v U_v(v_a)-C(a) \tag{18}$$

其中，ω_{rss} 和 ω_v 是效用函数相关属性的权重。U_{rss} 和

U_v 分别是 RSS 和速度的效用函数，C 是动作的开销函数。

对用户来说，RSS 越大，对应网络的链路质量越好，RSS 的效用函数 U_{rss} 可表示为[22]

$$U_{rss}(rss_a)=\eta_1\left(\frac{rss_a-rss_{\min}}{rss_{\min}}\right)^2 \tag{19}$$

其中，$rss_{\min}$ 是维持正常连接的最小 RSS，η_1 为归一化参数，用来维持 U_{rss} 的取值在[0,1]的范围内。

速度的效用函数 U_v 有 2 个组成部分，v_a 代表用户相对于基站的运行速度，可以表示用户接近或远离基站；v 代表用户的绝对速度，v 过高时，处于 ATC 小区的用户会面临频繁的切换，为改善这一问题，本文设定了一个速度阈值 V_T，当超过 V_T 时，用户将更倾向于选择卫星网。从而速度的效用函数可表示为[23]

$$U_v(v_a)=\begin{cases}\eta_2 v_a-\eta_3 vu(v-V_T), a\in ATC-subnetwork\\ \eta_4 vu(v-V_T), a\in GEO-subnetwork\end{cases} \tag{20}$$

其中，η_2 η_3 和 η_4 为归一化参数。

最后一个属性是网络开销函数，网络开销函数与切换开销以及用户电池寿命相关。对比 ATC 地面网络和 GEO 卫星网络，定义地面网开销为 1，卫星网开销为 $\psi(\psi>1)$

$$C(a)=\begin{cases}\psi, a\in ATC-subnetwork\\ 1, a\in GEO-subnetwork\end{cases} \tag{21}$$

3.2.3 Q 学习算法实现原理

Q 学习算法是一个在线学习算法，在算法执行过程中通过 2 阶段完成同步：1）探索阶段：学习者学习并获得最佳切换决策集；2）执行阶段：根据切换决策集执行切换。在每个决策时刻，学习者会以确定概率选择进行探索过程或者执行过程。当选择探索过程时，学习者随机选择切换至某个候选网络；选择执行过程时，学习者根据存储的 Q 值表选取 Q 值最高的网络，即根据现有策略集选取合适的动作。无论选择哪个过程，Q 值都会被更新并存储。如前文所述，当状态-动作对的访问次数 $T(s,a)$ 趋向于无穷时，学习率 ρ 趋向于 0，此时 $M_{qt}(s,a)$ 收敛至 $M_q^*(s,a)$。Q 学习算法在切换过程中的数学步骤如下[22]。

算法 1 Q 学习算法执行过程

初始化：

Q 值函数：$M_q(s,a)=r(s,a),\forall s\in S,a\in A$

历经的“状态-动作”数：

$T(s,a)=0,\forall s\in S,a\in A$

重复：

若处于探索阶段

随机选择动作：a_t

否则若处于执行阶段

令 $a_t=\max_{a\in A}[M_q(s_t,a)]$

$T(s_t,a_t)\leftarrow T(s_t,a_t)+1$

根据式（17）更新 $M_q(s_t,a_t)$

$s_t\leftarrow s_{t+1}$

$t\leftarrow t+1$

直到运行结束

算法 1 描述了切换决策的流程，在初始化阶段 Q 值函数被赋值为对应状态—动作对的回报函数。所谓执行阶段，即用户在切换时刻 t 按以下流程运作[22]。

1) 确定当前用户状态 s_t。收集邻近网络的 RSS 和用户速度，计算用户相对于基站的相对速度和下一时刻的 RSS，确定候选网络。

2) 做出决策。基于用户当前状态，在 Q 值表中查找现有最优决策，使用户在执行动作时 a_t 获得最大 Q 值：$a_t=\max_{a\in A}[M_q(s_t,a)]$。

3) 更新 Q 值函数。根据本次决策结果，通过式（16）更新 Q 值函数。

若在探索阶段，将步骤 2）换为随机选取一个网络即可。

4 仿真验证与分析

本研究的目标是让用户做出正确的决策，从而保证其正在进行业务的连续性。衡量切换性能的两大指标为切换掉线率和切换次数，本文对切换掉线率、切换次数和收益函数这几个参数进行了仿真分析[6,18,19,23]。仿真场景如图 1 所示，设置了一个 GEO 卫星波束，波束内有 3 个 ATC 小区重叠覆盖。假设 GEO 波束的半径为 1 500 km，ATC 小区的半径为 2 km，卫星链路建模为一个两状态信道模型，卫星仰角为 40°，RSS 的低磁滞门限为 30 dBm，决策时刻间隔为 0.6 s[19]。用户的平均连接时间服从指数分布，并将其平均值归一化为 1，则用户到达率服从泊松分布[6]。用户速度在 1~200 km/h 内变化，速度门限设定为 60 km/h。仿真中将本文所提方案与 2 种典型的切换方案进行比较，这 2 种方案是：1）基于 RSS 单门限的决策方

法（简称 RSS-T）[24]；2）基于 RSS 磁滞门限的方法（简称 RSS-H）[15]。

图 4 比较了以上几种方法在不同速度情况下的切换掉线率的变化情况。由于在高速场景下链路质量较差，切换掉线率随速度增加而变大，同等条件下，使用 RSS-T 和 RSS-H 方法更容易掉线。图 5 比较了在到达率不同的情况下切换掉线率的变化，结果表明切换掉线率随用户到达率线性增长，但是所提方法比其他方案约有 10%的改善。从图 4 和图 5 很容易看出本文所提方案优于对比方案，这是因为 RSS-T 和 RSS-H 都是由用户基于当前 RSS 进行切换决策，而所提方案是基于预测的下一时刻 RSS 做决策并提前执行切换，显然使用预测的方法能有效降低切换掉线率。

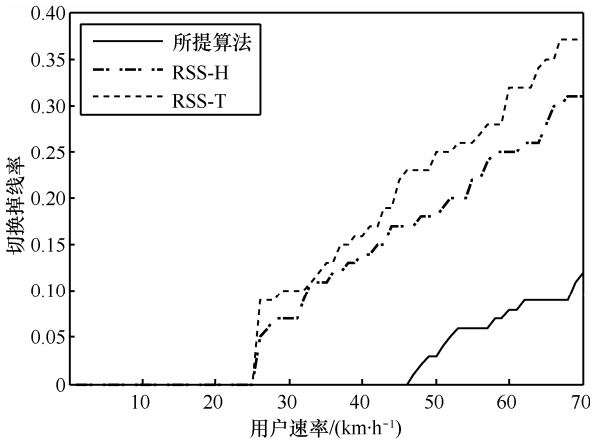


图 4　速率变化情况下切换掉线率的比较

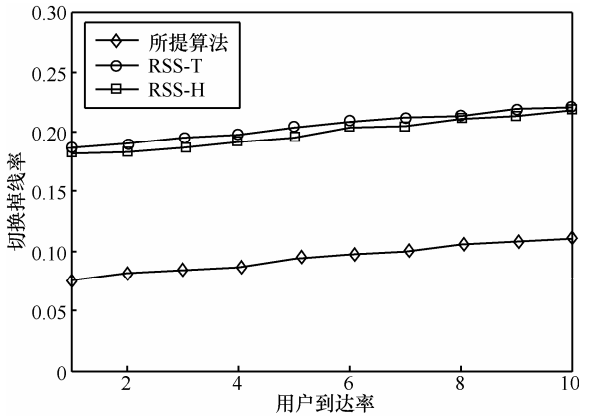


图 5　用户到达率变化情况下切换掉线率的比较

图 6 比较了 3 种方案在不同速度情况下的切换次数，结果表明所提方案具有最好的性能。RSS-H 方法可以有效改善乒乓效应，所以相较于 RSS-T，RSS-H 方案的性能更好。所提方案不仅用到了 RSS 预测，还用到了 Q 学习算法做切换决策，考虑到未来的收益，使切换决定更加准确。Q 学习中的收益函数还考虑到了用户速度，可以有效减少在高速场景下的切换次数，当速度超过门限值，根据式（20）可得，用户在 ATC 地面网络的收益减小，在卫星网络的收益变大，用户更倾向于切换至卫星网，所以图 6 中，当用户速度大于 60 km/h 的门限值时，切换次数不再增加。

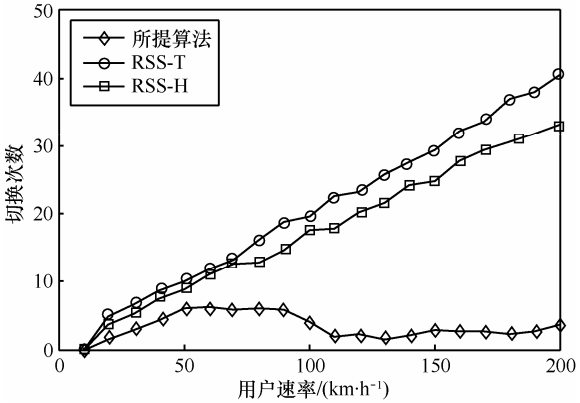


图 6　速率变化情况下切换次数的比较

图 7 比较了在不同到达率的情况下切换回报函数的变化情况。回报值由式（18）计算得来，可以看出，RSS-T 和 RSS-H 的回报值非常接近，而所提方案综合考虑了 RSS 和用户速度，以最大化长期收益为目标，具有更好的性能。

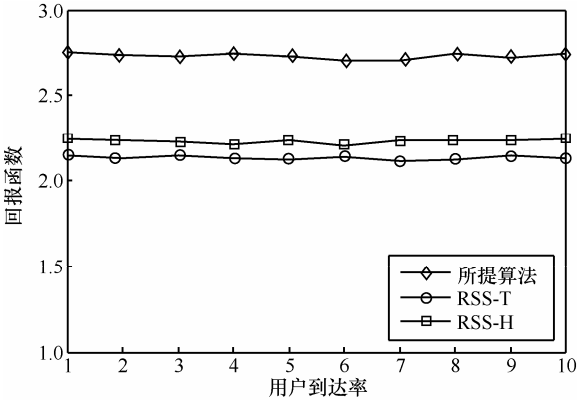


图 7　用户到达率变化情况下回报函数的比较

5　结束语

本文提出了一种基于 Q 学习和 RSS 预测的切换决策算法，RSS 预测使切换决策提前，能有效减少切换掉线率，降低切换次数，Q 学习算法时刻以最大化长期收益为目标，有效增加系统收益，并且适应高速运动场景。仿真结果验证了所提方案的正

确性和适用性，用户体验和切换准确度得到提高。在本文研究基础上，下一步将继续研究星地融合网络的接纳控制机制。

参考文献：

[1] CHINI P, GIAMBENE G, KOTA S. A survey on mobile satellite systems[J]. International Journal of Satellite Communications and Networking, 2010,28(1):29-57.

[2] AHN D S, KIM H W, AHN J, *et al*. Integrated/hybrid satellite and terrestrial networks for satellite imt-advanced services[J]. International Journal of Satellite Communications and Networking, 2011, 29(3):269-282.

[3] PARSONS, SINGH R. An ATC primer: the future of communications[J]. Mobile Satellite Ventures, 2006.

[4] HALVORSON E, EISENMAN A, EDALAT F, *et al.* Global resource manager for mobile satellite systems with ancillary terrestrial components[A]. IEEE Sarnoff Symposium[C]. 2010. 1-6.

[5] ZOU D, MENG W, HAN S. Euclidean distance based handoff algorithm for fingerprint positioning of WLAN system[A]. IEEE Wireless Communications and Networking Conference (WCNC)[C]. 2013. 1564-1568.

[6] CHANG B J, CHEN J F, HSIEH C H, *et al.* Markov decision process-based adaptive vertical handoff with RSS prediction in heterogeneous wireless networks[A]. IEEE Wireless Communications and Networking Conference(WCNC)[C]. 2009. 1-6.

[7] SODERMAN P, EKLUND J, GRINNEMO K J, *et al*. Handover in the wild: the feasibility of vertical handover in commodity smartphones[A]. IEEE International Conference on Communications(ICC)[C]. 2013. 6401-6406.

[8] LIU S, CHANG Y, WANG G, YANG D. Vertical handoff scheme concerning mobility in the two-hierarchy network[A]. IEEE GLOBECOM Workshops (GC Wkshps)[C]. 2011. 237-241.

[9] LAHBY M, CHERKAOUI L, ADIB A. An enhanced-topsis based network selection technique for next generation wireless networks[A]. International Conference on Telecommunications (ICT)[C]. 2013. 1-5.

[10] WU J S, YANG S F, HWANG B J. A terminal-controlled vertical handover decision scheme in IEEE 802.21-enabled heterogeneous wireless networks[J]. International Journal of Communication Systems, 2009,22(7):819-834.

[11] CHEN J, WEI Z, WANG Y, *et al*. A service-adaptive multi-criteria vertical handoff algorithm in heterogeneous wireless networks[A]. International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)[C]. 2012. 899–904.

[12] ZHANG X, LI Y, LI J, ZHAO K, ZHANG T. Proximate control stream assisted video transcoding for heterogeneous content delivery network[A]. IEEE International Conference on Image Processing (ICIP)[C]. 2014. 2552-2555.

[13] YIN R, CHAI R, CHEN Q. Joint utility optimization based vertical handoff algorithm in heterogeneous network[A]. IEEE Global Communications Conference (GLOBECOM)[C]. 2012. 5243-5248.

[14] WATKINS C, DAYAN P. Q-learning, technical note[J]. Machine Learning, 1992, 8: 279-292.

[15] MARICHAMY P, CHAKRABARTI S, MASKARA S L. Performance evaluation of handoff detection schemes[A]. Conference on Convergent Technologies for the Asia-Pacific Region (TENCON)[C]. 2003. 643-646.

[16] RISTIC B, ARULAMPALAM S, GORDON N. Beyond the Kalman Filter: Particle Filters for Tracking Applications[M]. Artech House, 2004.

[17] HARVEY, ANDREW C. Forecasting, Structural Time Series Models and the Kalman Filter[M]. Cambridge University Press, 1990.

[18] VEERAVALLI V V, KELLY O E. A locally optimal handoff algorithm for cellular communications[J]. IEEE Transactions on Vehicular Technology, 1997,146(3):603-609.

[19] SADEK M, ISSA S A. Handoff algorithm for mobile satellite systems with ancillary terrestrial component[A]. IEEE International Conference on Communications (ICC)[C]. 2012. 2763-2767.

[20] LIOLIS K P, OMEZ-VILARDEB O J G, CASINI E, *et al.* Statistical modeling of dual-polarized mimo land mobile satellite channels[J]. IEEE Transactions on Communications, 2010,58(11):3077-3083.

[21] CORAZZA G E, VATALARO F. A statistical model for land mobile satellite channels and its application to nongeostationary orbit systems[J]. IEEE Transactions on Vehicular Technology, 1994,43(3): 738-742.

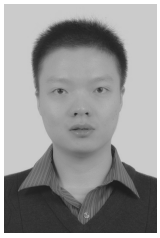
[22] TABRIZI H, FARHADI G, CIOFFI J. Dynamic handoff decision in heterogeneous wireless systems: Q-learning approach[A]. IEEE International Conference on Communications (ICC)[C]. 2012. 3217-3222.

[23] NIE J, HAYKIN S. A dynamic channel assignment policy through Q-learning[J]. IEEE Transactions on Neural Networks, 1999,10(6): 1443-1455.

[24] POLLINI G P. Trends in handover design[J]. IEEE Communications Magazine, 1996,34(3):82-90.

作者简介：

熊丹妮（1990-），女，湖北武汉人，北京邮电大学硕士生，主要研究方向为空间通信、无线通信网络、多种网络间切换技术等。



李屹（1977-），男，浙江东阳人，博士，北京邮电大学副教授，主要研究方向为未来网络、5G 通信、星地融合通信等。