基于强化学习的5G网络切片虚拟网络功能迁移算法

唐 伦 周 钰* 谭 颀 魏延南 陈前斌
(重庆邮电大学通信与信息工程学院 重庆 400065)
(重庆邮电大学移动通信重点实验室 重庆 400065)

摘 要：针对5G网络切片架构下业务请求动态性引起的虚拟网络功能(VNF)迁移优化问题，该文首先建立基于受限马尔可夫决策过程(CMDP)的随机优化模型以实现多类型服务功能链(SFC)的动态部署，该模型以最小化通用服务器平均运行能耗为目标，同时受限于各切片平均时延约束以及平均缓存、带宽资源消耗约束。其次，为了克服优化模型中难以准确掌握系统状态转移概率及状态空间过大的问题，该文提出了一种基于强化学习框架的VNF智能迁移学习算法，该算法通过卷积神经网络(CNN)来近似行为值函数，从而在每个离散的时隙内根据当前系统状态为每个网络切片制定合适的VNF迁移策略及CPU资源分配方案。仿真结果表明，所提算法在有效地满足各切片QoS需求的同时，降低了基础设施的平均能耗。

关键词：5G网络切片；虚拟网络功能迁移；强化学习；资源分配
中图分类号：TN929.5 **文献标识码**：A **文章编号**：1009-5896(2020)03-0669-09
DOI: 10.11999/JEIT190290

Virtual Network Function Migration Algorithm Based on Reinforcement Learning for 5G Network Slicing

TANG Lun ZHOU Yu TAN Qi WEI Yannan CHEN Qianbin
(*School of Communication and Information Engineering*, *Chongqing University of Post and Telecommunications*, *Chongqing* 400065, *China*)
(*Key Laboratory of Mobile Communication Technology*, *Chongqing University of Post and Telecommunications*, *Chongqing* 400065, *China*)

Abstract: In order to solve the Virtual Network Function (VNF) migration optimization problem caused by the dynamicity of service requests on the 5G network slicing architecture, firstly, a stochastic optimization model based on Constrained Markov Decision Process (CMDP) is established to realize the dynamic deployment of multi-type Service Function Chaining (SFC). This model aims to minimize the average sum operating energy consumption of general servers, and is subject to the average delay constraint for each slicing as well as the average cache, bandwidth resource consumption constraints. Secondly, in order to overcome the issue of having difficulties in acquiring the accurate transition probabilities of the system states and the excessive state space in the optimization model, a VNF intelligent migration learning algorithm based on reinforcement learning framework is proposed. The algorithm approximates the behavior value function by Convolutional Neural Network (CNN), so as to formulate a suitable VNF migration strategy and CPU resource allocation scheme for each network slicing according to the current system state in each discrete time slot. The simulation results show that the proposed algorithm can effectively meet the QoS requirements of each slice while reducing the average energy consumption of the infrastructure.

Key words: 5G network slicing; Virtual Network Function (VNF) migration; Reinforcement learning; Resource allocation

收稿日期：2019-04-25；改回日期：2019-09-11；网络出版：2019-09-19
*通信作者： 周钰 137068966@qq.com
基金项目：国家自然科学基金(61571073)，重庆市教委科学技术研究项目(KJZD-M201800601)
Foundation Items: The National Natural Science Foundation of China (61571073), The Science and Technology Research Program of Chongqing Municipal Education Commission (KJZD-M201800601)

1 引言

未来5G网络在系统架构和性能优化等多方面将进一步改进[1]。借助网络切片及网络功能虚拟化技术(Network Function Virtualization, NFV)，使得服务功能链(Service Function Chaining, SFC)在基础设施中的移动性管理成为了可能。随着网络负载及基础设施中节点与链路状态的变化，对SFC中虚拟网络功能(Virtual Network Function, VNF)进行迁移，可以在提高底层资源利用率的同时满足不同切片对时延的要求。此外，VNF的灵活编排也为节约系统能耗提供了有利条件，在VNF共享的机制下，对资源利用率较低服务器上的VNF进行迁移，并关闭相应的服务器可以达到降低能耗的目的[2]。

目前为止，大多数文献如文献[3,4]研究了在一定迁移触发时机下对虚拟机进行迁移的相关问题，然而这类迁移策略并没有同时考虑多条SFC的业务场景。文献[5]研究的是服务功能链VNF与节点VNF实例呈一一映射关系下的VNF迁移问题，针对VNF实例共享状态下，若某个切片性能无法满足用户需求，目前的研究无法制定有效的策略实现VNF的迁移。文献[6]提出了基于MDP理论的VNF迁移算法以应对不断变化的工作负载，其迁移策略最大限度地降低能耗以及VNF迁移造成的重配置成本为目标。文献[7]建立一种成本模型并提出贪婪算法优化VNF的迁移，然而该方案只能解决单一调度周期上的资源分配问题。

综上所述，目前大多数的工作停留在固定网络环境下瞬时指标优化的问题上，而针对具备一定生命周期的动态切片业务场景，研究利用在线的资源管理技术优化VNF迁移在长时间尺度下平均性能指标的工作不多。基于业务请求动态变化的5G网络切片场景，本文联合考虑了CPU、带宽及缓存资源，并将VNF的动态迁移及资源分配看作是一个无穷时间马尔可夫决策过程，本文的主要贡献有：(1) 设计基于流量感知的VNF队列模型，将不同种类型SFC的VNF部署过程模拟成M/M/1排队过程，实现了VNF实例的共享；(2) 建立了基于受限马尔可夫决策过程(Constrained Markov Decision Process, CMDP)的随机优化模型以实现多类型SFC的动态部署，本文模型以最小化通用服务器平均运行能耗为目标，同时受限于各切片平均时延约束以及平均缓存、带宽资源消耗约束；(3)提出了一种基于深度Q学习框架的VNF智能迁移学习算法，本文方法引入经验回放池及目标Q网络，通过卷积神经网络近似行为值函数，从而在每个离散的时隙内根据当前系统状态为每个网络切片制定合适的VNF迁移策略及CPU资源分配方案。

2 基于能耗优化的VNF队列模型

2.1 网络模型

本文主要研究的是长时间尺度下基于5G网络切片场景中多类型SFC的VNF迁移优化问题。系统场景如图1所示，其包含3层，应用层主要负责为每

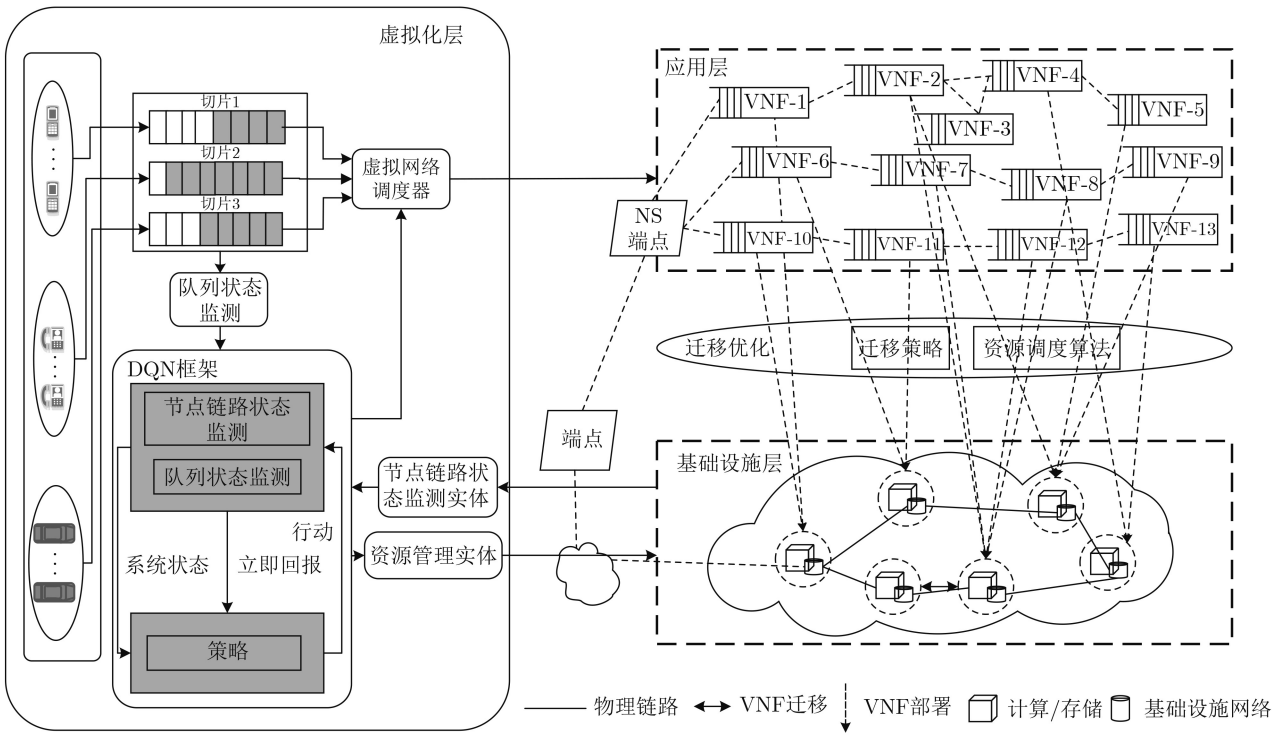

图 1 5G网络切片架构下的VNF迁移系统场景图

个切片请求提供有序的VNF集合来处理到达的数据流，基础设施层中云服务器提供包含计算资源、缓存资源、带宽资源等多种类型的虚拟网络资源，虚拟化层根据虚拟网络用户的业务状态、QoS需求等实现VNF的动态迁移、虚拟网络资源的灵活分配。

本文将切片定义为不同用户对同一类型网络业务请求的数据集合，用$\boldsymbol{S}=\{S_1,S_2,\cdots,S_I\}$表示，其中$I$为整个网络服务的切片业务种类数。实现切片业务的SFC由一组有序的VNF组成，定义网络中VNF种类的集合用$\boldsymbol{F}=\{f_1,f_2,\cdots,f_J\}$表示，并假设通用服务器节点$h\in H$上实例化的VNF集合为$\boldsymbol{F}_h$。对于到达节点每个VNF实例$f_j^s\in\boldsymbol{F}$的数据流，都存在一个M/M/1的排队过程，假设切片S_i的数据包到达率服从均值为$\overline{\lambda}_i$且在时隙间独立的某种分布，并令时隙t切片S_i的VNFf_{ij}数据包到达率为$\lambda_{ij}(t)$。在处理数据包的过程中，若实现切片S_i的初始VNF为f_j，则令$P(f_j|o,i)=1$，否则为0；令$P(f_2|f_1,i)$表示用户请求切片S_i的数据流在VNFf_1被处理后传输至VNFf_2的比例，若$P(f_2|f_1,i)=0$，则说明在实现切片S_i的SFC中，VNF模块f_1的下一个VNF模块不是VNFf_2。

虚拟网络由$G^s=(V^s,E^s)$表示，其中V^s表示虚拟节点的集合，E^s表示虚拟链路的集合。对于节点h上的VNF实例$f_{h,j}^s$，其服务速率$\mu_{f_{h,j}^s}$由节点分配给该VNF实例的CPU资源大小所决定，令$\mu_{f_{h,j}^s}=Z_{f_{h,j}^s}\cdot\xi$，其中$Z_{f_{h,j}^s}$表示节点$h$分配给VNF实例$f_{h,j}^s$的CPU资源，$\xi$表示服务速率系数，该系数表示VNF提供一定的服务速率所对应的CPU资源关系[8]。任意节点$h\in H$具有的固定CPU大小为κ_h，节点之间的传输时延用$\delta(h,l)$表示，由于设备故障或负载过大等原因，节点之间的通信链路存在失效的可能，令$\eta_{h,l}=1$表示节点h与l之间的链路处于正常状态，否则失效；$\zeta_h=1$表示节点h运行正常，否则失效。

令$B(h,f)\in\{0,1\}$表示VNF$f\in F$是否部署在节点h，且满足$\sum_{h\in H}B(h,f)=1,\forall f\in F$。定义辅助变量$\hat{\lambda}_{ij}$表示切片$S_i$中进入VNF$f_j$的总数据量，且对于任何$i\in I$都满足

$$\hat{\lambda}_{ij}=P(f_j|o,i)\lambda_{ij}+\sum_{f_p\in F\backslash\{f_j\}}P(f_j|f_p,i)\hat{\lambda}_{ip}\tag{1}$$

因此对于节点h的VNF实例$f_{h,j}^s$，到达的总数据量可以表示为$\phi_{f_{h,j}^s}=\sum_{i\in I}\hat{\lambda}_{ij}B(h,f_{ij})$。节点$h$剩余CPU资源表示为$\vartheta_h=\kappa_h-\sum_{f_j^s\in F_h}Z_{f_{h,j}^s}$，令VNF$f_{ij}$的处理时延用$D_{\mathrm{pro}}(ij)$表示，可以得出

$$D_{\mathrm{pro}}(ij)=\sum_{h\in H}\left(\frac{Q_{h,j}}{\overline{\lambda}_{f_{h,j}^s}}+\frac{\bar{P}}{\mu_{f_{h,j}^s}}\right)\cdot B(h,f_{ij}),\ \ \forall f\in F\tag{2}$$

其中，$Q_{h,j}$表示节点h的VNF实例$f_{h,j}^s$当前的队列长度大小，$\overline{\lambda}_{f_{h,j}^s}=\mathrm{E}[\phi_{f_{h,j}^s}(t)]$为$f_{h,j}^s$的数据包达到过程的均值，并假设数据包大小服从均值为$\bar{P}$的指数分布[9]。节点h的VNF实例$f_{h,j}^s$队列更新过程可以表示为

$$Q_{h,j}(t+1)=\max\left[Q_{h,j}(t)-\mu_{f_{h,j}^s}(t),0\right]+\phi_{f_{h,j}^s}(t)\tag{3}$$

为了计算数据流在链路传输产生的时延，先确定网络切片i调度VNF$f_j\in F$的期望平均次数σ_{ij}，即

$$\sigma_{ij}=P(f_j|o,i)+\sum_{f_p\in F\backslash\{f_j\}}P(f_j|f_p,i)\sigma_{ip}\tag{4}$$

其中，第1项表示实现切片S_i的SFC初始VNF是否为f_j，第2项表示VNFf_p的数据进入VNFf_j的比例。通过得到的σ_{ij}，可以计算网络切片i的传输时延为

$$D_{\mathrm{tr}}(i)=\sum_{f_j,f_k\in F_{S_i}}\sigma_{ij}P(f_k|f_j,i)\cdot\sum_{h,l\in H}\delta(h,l)B(h,f_{ij})B(l,f_{ik})\tag{5}$$

因此网络切片i的VNF调度总时延可以表示为

$$D_i=\sum_{f_j\in\boldsymbol{F}_{s_i}}D_{\mathrm{pro}}(ij)+\sum_{f_j,f_k\in\boldsymbol{F}_{S_i}}\sigma_{ij}P(f_k|f_j,i)\cdot\sum_{h,l\in H}\delta(h,l)B(h,f_{ij})B(l,f_{ik})\tag{6}$$

其中，$\boldsymbol{F}_{S_i}$表示组成网络切片S_i的所有VNF集合。令L_{ij}表示f_{ij}从VNF实例$f_{h,j}^s$队列离开的数据包，满足

$$L_{ij}=\min[\mu_{f_{h,j}^s},Q_{h,j}]\cdot\frac{Q_{f_{ij}}}{Q_{h,j}}\tag{7}$$

其中，$Q_{f_{ij}}$表示在VNF实例$f_{h,j}^s$的队列$Q_{h,j}$中包含切片S_i的数据量。令$X_{h,l}$表示节点h至节点l的链路带宽消耗，可以表示为

$$X_{h,l}=\sum_{i\in I}\sum_{f_j,f_k\in\boldsymbol{F}_{S_i}}\overline{P}L_{ij}(f_k|f_j,i)B(f_{ij},h)B(f_{ik},l)\tag{8}$$

定义节点h的负载密度ρ_h为当前VNF实例的总负载与能够提供的最大服务速率之比，即表示为

$$\rho_h=\begin{cases}\displaystyle\sum_{f_{h,j}^s\in F_h}\left(\phi_{f_{h,j}^s}+Q_{h,j}\right)/\mu_h,\\ \displaystyle\sum_{f_{h,j}^s\in F_h}\left(\phi_{f_{h,j}^s}+Q_{h,j}\right)<\mu_h\\ 1,\qquad\text{其他}\end{cases}\tag{9}$$

其中，μ_h表示节点h能够提供的最大服务速率。本文所优化的能耗模型主要包含两部分内容：节点处于开启状态时产生的基础能耗和随着VNF实例负载变化产生的运行能耗[10]。具体可以表示为

$$C(\rho, H_{\text{on}}) = \sum_{h\in H}[(1-u_h)\rho_h P_h + \varsigma_h^t u_h P_h] \tag{10}$$

其中，$\varsigma_h^t = 1$表示在时刻t节点h处于开启的状态，否则为0; $u_h \in (0,1)$为节点h恒定能耗百分比；P_h表示节点h的CPU资源均被占用产生的最大能耗。

2.2 问题描述

为了表述队列Q在时间域上负载的变化，本文将VNF的迁移及CPU资源的分配描述为一个受限的马尔可夫决策问题(CMDP)，即定义一个4元组$M =< R, A, \mathcal{P}, \mathcal{C} >$，其基本元素包括系统状态、迁移行为、状态转移概率和成本函数。定义系统在时刻t的状态为$r(t) = (\boldsymbol{Q}(\boldsymbol{t}), \zeta(\boldsymbol{t}), \boldsymbol{\eta}(\boldsymbol{t})) \in R$，定义系统在时刻$t$的行动为$a(t) = \{\boldsymbol{\Psi}(\boldsymbol{t}), \boldsymbol{Z}(\boldsymbol{t})\} \in A$。其中，$\boldsymbol{\Psi}(\boldsymbol{t})$为时隙$t$内的VNF迁移的动作向量，其元素为$\psi_{f_{ij}}(t)$。决策由两部分组成，包括$\varphi_{f_{ij}}(t), h^*_{f_{ij}}(t)$, $\forall f_{ij} \in F, i \in I$, $\varphi_{f_{ij}}(t) = 1$表示在时隙t对切片i的VNFf_j进行迁移，否则为0；$h^*_{f_{ij}}(t)$表示VNFf_{ij}迁移的目标节点；$\boldsymbol{Z}(\boldsymbol{t})$为时隙$t$内每个VNF的CPU资源分配动作集合。

$\mathcal{P}(r(t), a(t), s(t+1))$表示状态转移概率，令$\pi : R \to A$代表一个稳定的确定性策略，其将状态空间映射到行动空间上，即$a = \pi(r)$。在时隙t根据策略$\pi \in \Pi$，则期望累积折扣回报可以表示为

$$\bar{C}^{\pi}(r) = \mathrm{E}_{\pi}\left[\sum_{t=0}^{\infty}\gamma^t C(r_t, \pi(r_t))|r_0 = r\right] \tag{11}$$

其中$C(r_t, \pi(r_t))$由式(10)可得，期望累积折扣端到端时延可以表示为

$$\bar{D}_i^{\pi}(r) = \mathrm{E}_{\pi}\left[\sum_{t=0}^{\infty}\gamma^t D_i(t)|r_0 = r\right], \forall i \in I \tag{12}$$

期望累积折扣缓存资源消耗可以表示为

$$\begin{aligned}\bar{N}_h^{\pi}(r) &= \mathrm{E}_{\pi}\left[\sum_{t=0}^{\infty}\gamma^t N_h(t)|r_0 = r\right] \\ &= \mathrm{E}_{\pi}\left[\sum_{t=0}^{\infty}\gamma^t \sum_{f^s_{h,j}\in F_h}\left(\phi_{f^s_{h,j}} + Q_{h,j}\right)|r_0 = r\right], \\ &\quad \forall h \in H\end{aligned} \tag{13}$$

期望累积折扣带宽资源消耗可以表示为

$$\bar{X}_{h,l}^{\pi}(r) = \mathrm{E}_{\pi}\left[\sum_{t=0}^{\infty}\gamma^t X_{h,l}(t)|r_0 = r\right], \forall h \in H \tag{14}$$

其中，$\gamma \in (0,1]$为折扣因子，$D_i(t)$由式(6)可得，$X_{h,l}(t)$由式(8)可得。本文的目标是找到最优的VNF迁移决策$\boldsymbol{\Psi}(\boldsymbol{t})$及资源分配策略$\boldsymbol{Z}(\boldsymbol{t})$从而保证QoS的同时实现能耗的优化，进而随机优化模型可以表示为

$$\left.\begin{aligned}&\min_{\boldsymbol{\Psi}(\boldsymbol{t}),\boldsymbol{Z}(\boldsymbol{t})} \bar{C}^{\pi}(r) \\ &\text{s.t.}\ \ \mathrm{C1}: \bar{D}_i^{\pi}(r) \le \tau_i, \forall i \in I \\ &\qquad \mathrm{C2}: \bar{N}_h^{\pi}(r) \le \chi_h, \forall h \in H \\ &\qquad \mathrm{C3}: \bar{X}_{h,l}^{\pi}(r) \le \Delta_{h,l}, \forall h,l \in H \\ &\qquad \mathrm{C4}: Z_h^f(t) \ge 0, \forall h \in H \\ &\qquad\quad \sum_{h\in H} Z_h^f(t) \le \kappa_h\end{aligned}\right\} \tag{15}$$

其中，C1中$\tau_i, \forall i \in I$为各个切片的时延约束，C2中$\chi_h, \forall h \in H$为各主机缓存资源消耗约束，C3中$\Delta_{h,l}$, $\forall h,l \in H$为链路带宽资源消耗约束，C4表示每个主机分配个VNF的CPU资源不少于0，且分配给VNF的CPU总资源不超过每台主机固定的CPU容量。

2.3 问题转换

式(15)CMDP问题可以通过拉格朗日理论转化为不受限MDP问题，定义式(15)对应的拉格朗日函数为

$$\begin{aligned}L(\boldsymbol{\beta}, r, \pi) = \mathrm{E}_{\pi}&\left\{\sum_{t=0}^{\infty}\gamma^t g^{\boldsymbol{\beta}}(r(t), a(t))|r_0 = r\right\} \\ &- \sum_{i\in I}\beta_i^d\delta_i - \sum_{h\in H}\beta_h^q\chi_h - \sum_{h,l\in H}\beta_{h,l}^x\Delta_{h,l}\end{aligned} \tag{16}$$

其中，$g^{\boldsymbol{\beta}}(r(t), a(t)) = C(t) + \sum_{i\in I}\beta_i^d D_i(t) + \sum_{h\in H}\beta_h^q \cdot \sum_{f^s_{h,j}\in F_h}\phi_{f^s_{h,j}}(t) + \sum_{h,l\in H}\beta_{h,l}^x X_{h,l}(t)$为时隙$t$的拉格朗日回报，$\boldsymbol{\beta} : \beta \ge 0$为拉格朗日乘子。定义状态值函数为

$$V^{\pi,\boldsymbol{\beta}}(r) = \mathrm{E}_{\pi}\left\{\sum_{t=0}^{\infty}\gamma^t g^{\beta}(r(t), \pi(r(t)))|r_0 = r\right\} \tag{17}$$

因此，优化问题式(15)可转化为式(18)的无约束MDP问题

$$\begin{aligned}\min_{\pi\in\Pi}\max_{\boldsymbol{\beta}\ge 0}&\left(V^{\pi,\beta}(r) - \sum_{i\in I}\beta_i^d\delta_i - \sum_{h\in H}\beta_h^q\chi_h\right. \\ &\left. - \sum_{h,l\in H}\beta_{h,l}^x\Delta_{h,l}\right)\end{aligned} \tag{18}$$

其对偶问题为

$$\begin{aligned}\max_{\boldsymbol{\beta}\ge 0}\min_{\pi\in\Pi}&\left(V^{\pi,\beta}(r) - \sum_{i\in I}\beta_i^d\delta_i - \sum_{h\in H}\beta_h^q\chi_h\right. \\ &\left. - \sum_{h,l\in H}\beta_{h,l}^x\Delta_{h,l}\right)\end{aligned} \tag{19}$$

对于给定的向量$\boldsymbol{\beta}:\beta\geq 0$，无约束问题式(19)对应的最优策略应满足贝尔曼最优性方程为

$$V^{*,\boldsymbol{\beta}}(r)=\min_{a\in A}\left(g^{\boldsymbol{\beta}}(r,a)+\gamma\sum_{r'\in R}\mathcal{P}\left(r'|r,a\right)V^{*,\boldsymbol{\beta}}\left(r'\right)\right) \tag{20}$$

其中，$V^{*,\boldsymbol{\beta}}:R\to C$称为最优状态值函数，令$Q^{*,\boldsymbol{\beta}}:R\times A\to C$表示最优行动值函数，其满足

$$Q^{*,\boldsymbol{\beta}}(r,a)=g^{\boldsymbol{\beta}}(r,a)+\gamma\sum_{r'\in R}\mathcal{P}(r'|r,a)V^{*,\boldsymbol{\beta}}(r') \tag{21}$$

因此，式(20)可以重写为

$$V^{*,\boldsymbol{\beta}}(r)=\min_{a\in A}Q^{*,\boldsymbol{\beta}}(r,a) \tag{22}$$

当$T\to\infty$时一定存在一个稳定的最优策略使目标函数最小。因此最优策略$\pi^{*,\boldsymbol{\beta}}$可以表示为

$$\pi^{*,\boldsymbol{\beta}}=\arg\min_{a\in A}Q^{*,\boldsymbol{\beta}}(r,a),\forall r\in R \tag{23}$$

3　基于深度Q学习(DQN)的VNF迁移优化算法

由上一节可知，式(23)中不同状态的最优动作最终组成了策略$\pi^{*,\boldsymbol{\beta}}$，由于本文数据包的到达、节点与链路失效的可能性都是随机的，无法通过值函数的贝尔曼迭代方式获得最优策略；同时，目前的相关研究缺少高效的机制确定$\varphi^{*}_{f_{ij}}(t)$与$h^{*}_{f_{ij}}(t),\forall f_{ij}\in F,i\in I$从而满足$[\boldsymbol{\Psi}^{*}(\boldsymbol{t}),\boldsymbol{Z}^{*}(\boldsymbol{t})]=\underset{\boldsymbol{\Psi}(\boldsymbol{t}),\boldsymbol{Z}(\boldsymbol{t})}{\arg\min}\,\bar{C}^{\pi}(r)$。本文接下来提出一种基于深度Q学习的VNF迁移优化算法对问题进行求解。本算法通过与神经网络相结合，能有效地解决状态与行动空间过大的问题，提高了算法学习的性能[11,12]。其本质上是利用神经网络训练函数f_{ap}来近似Q值的分布，该方法把状态r作为输入，经过神经网络分析后输出对应动作的Q值，即

$$\boldsymbol{Q}(r,a)\approx f_{\mathrm{ap}}(r,a,w) \tag{24}$$

其中，w为神经网络的权重，$\boldsymbol{Q}(r,a)=[Q(r,a_1),Q(r,a_2),\cdots,Q(r,a_n)]$。本文采用的神经网络模型为多层卷积神经网络(CNN)，网络结构主要包括输入层、卷积层、以及全连接层[13]，其中，输入层的原始输入包括全局队列状态、节点状态以及链路状态，然后通过卷积层进行局部特征提取。

DQN在Q网络的基础上增加了一个目标Q网络(fixed Q-targets)来计算目标Q值，目标Q网络在一段时间内才会更新一次，而Q网络在每次迭代过程之后都会更新。进一步，目标Q值y表示为

$$y=g^{\beta}(r,a)+\gamma\left[\min_{a'\in A}\boldsymbol{Q}(r',a',w^{-})\right] \tag{25}$$

其中，w^{-}为目标Q网络的权重。为了提高网络预测的性能，权重函数需要反复的学习与训练以拟合复杂的环境特征，具体的训练模型如图2所示。

该过程通过最小化Q网络和目标Q网络之间的损失函数来优化权重w，损失函数可以表示为

$$G(w)=\mathrm{E}\left[\left(y-\boldsymbol{Q}(r,a,w)\right)^{2}\right] \tag{26}$$

此外，DQN网络还运用经验回放池(experience replay)存放每次迭代的数据样本，并采用随机的方式从经验回放池中抽取一部分数据用于网络参数的更新，以此来打破数据间的关联。具体的学习流程如表1所示。

表1中，将原式(19)的对偶问题解耦成外层循环和内层循环两个部分，内层循环主要通过DQN网络学习的方式获得最优的VNF迁移方案，外层循环主要解决拉格朗日乘子的更新，本文利用随机次梯度法在线学习拉格朗日乘子$\boldsymbol{\beta}:\beta\geq 0$的最优值。可以证明，当$t\to\infty$时可满足$\beta$收敛，即得到原问题的最优解。

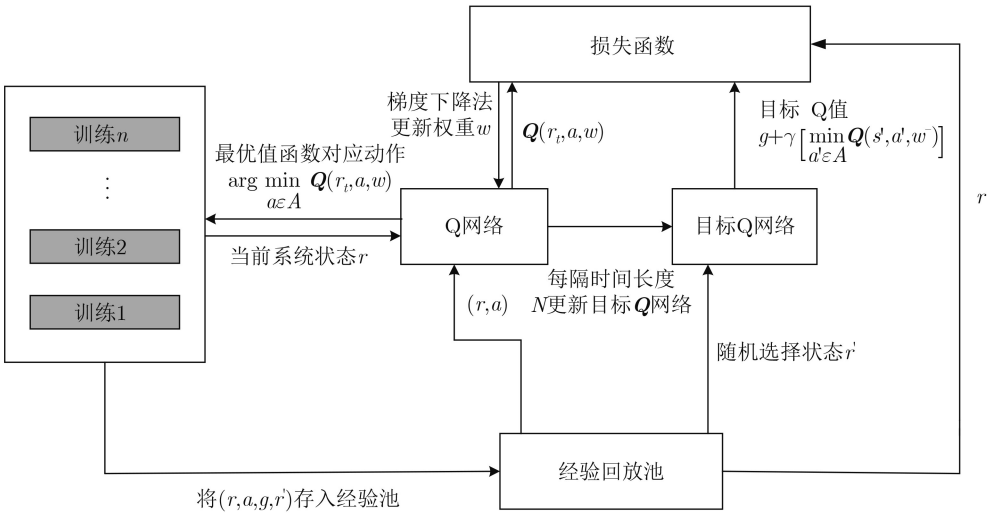

图 2 基于DQN的虚拟网络功能智能迁移学习架构图

通过对DQN网络的不断训练，可以更加精确地近似Q函数的分布，进而利用最新的网络求得更优的VNF迁移方案。具体的流程如表2所示。其中，步骤(3)是对网络状态进行监测，并将当前网络状态作为Q网络的输入。在步骤(4)—步骤(8)中，如果节点或链路在时隙t失效，为了提高网络的可靠性，系统在对相应的VNF进行迁移的基础上，选择最优的VNF迁移策略a_t^*对能耗与时延进行优化，否则直接选择最优的迁移策略a_t^*。

4 仿真结果与分析

本节将切片的平均端到端时延、通用服务器平均总功耗等作为评价指标，对所提算法进行了仿真验证分析。为了更好地评估本文基于DQN的VNF迁移算法的有效性，对比了Q学习(Q-learning)，文献[15]中的OVMP算法，以及文献[16]基于遗传算法(GA)的VNF动态部署方案。其中OVMP是一种基于阈值的迁移整合算法，该算法对计算资源利用率低于阈值的节点上运行的VNF迁移，并关闭相应的服务器以节约能耗。在遗传算法中，满足时延及资源约束的前提下，根据当前的网络状态，基于功耗最优化的策略对VNF进行迁移。仿真工具使用的是Ubuntu 18.04 LTS操作系统下的Python 3.0及TensorFlow框架。

表 1 基于DQN的价值函数近似

(1) 初始化Q网络，采用Xavier[14]初始化权重，即令权重的概率分布函数服从$W\sim U\left[-\frac{\sqrt{6}}{\sqrt{\upsilon_l+\upsilon_{l+1}}},\frac{\sqrt{6}}{\sqrt{\upsilon_l+\upsilon_{l+1}}}\right]$的均匀分布，初始化目标Q网络，权重为$w^-=w$，其中$l$为网络层数，$\upsilon$为神经元个数

(2) 初始化拉格朗日乘子$\beta_i^d\leftarrow 0,\beta_h^q\leftarrow 0,\beta_{h,l}^x\leftarrow 0,\forall i\in I,\forall h,l\in H$，初始化经验回放池

(3) for episode $k=1,2,\cdots,K$ do

(4) 随机选取一个状态初始化r_1

(5) for $t=1,2,\cdots,T$ do

(6) 随机选择一个概率p，if $p\geq\varepsilon$

(7) 计算VNF迁移及CPU资源分配策略$a_t^*=\arg\min\limits_{a\in A}\boldsymbol{Q}(r_t,a,w)$

(8) else 选择一个随机的行动$a_t\neq a_t^*$

(9) 执行行动a_t，获得拉格朗日回报$g^\beta(r_t,a_t)$，并观察下一时刻状态r_{t+1}

(10) 将经验样本$\left(r_t,a_t,g^\beta(r_t,a_t),r_{t+1}\right)$存入经验回放池中

(11) 从经验池中随机抽取一组Mini-batch的经验样本$\left(r_k,a_k,g^\beta(r_k,a_k),r_{k+1}\right)$

(12) 利用目标Q网络得到$\min\limits_{a'\in A}\boldsymbol{Q}(r_{t+1},a',w^-)$，求得$y_k=g^\beta(r_k,a_k)+\gamma\min\limits_{a'\in A}\boldsymbol{Q}(r_{t+1},a',w^-)$

(13) 对$(y_k-\boldsymbol{Q}(r_t,a_k,w))^2$使用梯度下降法对$w$进行更新

(14) 每隔时间长度T_q更新目标Q网络，即$w^-=w$

(15) 利用随机次梯度法更新拉格朗日乘子$\boldsymbol{\beta}:\beta\geq 0$

(16) end for

(17) end for

表 2 基于DQN的VNF在线迁移算法

(1) for $t=1,2,\cdots,T$ **do**

(2) *网络状态的监测*\

(3) 监测当前时隙t下的全局状态$r(t)$，包括全局队列状态$\boldsymbol{Q(t)}$、全局节点状态$\boldsymbol{\zeta(t)}$以及全局链路状态$\boldsymbol{\eta(t)}$

(4) if $\zeta_h(t)=0$或$\eta_{h,l}(t)=0$

(5) 在将满足$B(h,f)=1$或$P(f_p|f_j)B(f_j,h)B(f_p,l)\neq 0$的所有$\forall f\in F$迁移至其它节点的基础上，计算最优的VNF迁移策略及CPU资源分配策略$a_t^*=\arg\min\limits_{a\in A}\boldsymbol{Q}(r_t,a,w)$

(6) else

(7) 直接计算最优的VNF迁移策略及CPU资源分配策略$a_t^*=\arg\min\limits_{a\in A}\boldsymbol{Q}(r_t,a,w)$

(8) 基于最优行动a_t^*执行VNF的迁移，并进行资源的分配

(9) $t=t+1$

(10) end for

4.1 参数设置

本文假设的网络场景为全连接型网络，网络包含了8台通用服务器，假设总共存在8种不同类型VNF，每一台服务器实例化5种不同类型的VNF。进一步，本文考虑了3种对时延要求不同的网络切片业务，实现网络切片业务的SFC由3～8个不等的有序VNF组成，具体网络场景下的相关参数如表3所示。

表 3　仿真参数

仿真参数	仿真值	仿真参数	仿真值
网络切片业务数量I	3	服务器总台数H	8
VNF种类J	10	节点失效率	服从均值为[0.01,0.02]均匀分布
时隙长度T_s	10 s	链路失效率	服从均值为[0.02,0.04]均匀分布
数据包到达过程	独立同分布的泊松过程	链路传输时延δ	0.5 ms
平均数据包大小$\overline{P}$	500 kbit/packet	服务器最高功率P_h	800 W
节点缓存空间χ	300 MB	服务器功耗百分比u_h	0.3
节点CPU个数κ	8	最大迭代轮数	2000
单个CPU最大服务速率ξ	25 MB/s	总训练步长	200000
链路带宽容量Δ	640 Mbps	学习率α	0.0001
折扣因子γ	0.9	Mini-batch	8

本文DQN框架中的Q网络与目标Q网络采用的是包含3层卷积层、2层全连接层的多层卷积CNN网络，每一层的相关信息包括卷积核大小、卷积步长、卷积核个数等，具体的CNN网络参数设置如表4所示。进一步，在表1的相关参数设置中，Q网络权重的更新采用的是Adam优化器，并每经过$T_q = 200$次迭代对目标Q网络的权重进行一次更新，DQN经验池的容量设置为10000，贪婪策略中的$\varepsilon = 0.7$。

4.2 仿真结果分析

图3与图4分别描述了在本文所提基于DQN的VNF动态迁移算法下，各个切片数据包平均时延、缓存资源和链路带宽资源平均利用率随迭代轮次的变化情况。进一步，切片业务数据包的平均时延约束分别为20 ms, 30 ms以及40 ms，各个切片业务数据包达到率分别服从均值为4000packets/slot, 5000packets/slot以及6000packets/slot的泊松过程。从图3与图4中可以分别看出，各切片业务数据包的平均时延和资源平均利用率均随着迭代次数的增加逐渐增大并最终在对应约束附近收敛，这主要是因为在迭代初期，平均时延处于较低水平是以资源利用率较低和牺牲能耗为代价的，随着迭代次数的增加，系统在满足各个切片业务数据包的平均时延约束、缓存和链路带宽资源约束的同时，还优化了通用服务器产生的能耗，这也体现了所涉及的VNF迁移算法的有效性。

表 4　CNN神经网络参数

网络层	卷积核大小	卷积步长	卷积核个数	激活函数
卷积层1	7×7	2	32	ReLU
卷积层2	5×5	2	64	ReLU
卷积层3	3×3	1	64	ReLU
全连接层1	–	–	512	ReLU
全连接层2	–	–	122	Linear

图5与图6分别表示在不同时延约束下，本文所提基于DQN的VNF动态迁移算法与其它3种对比算法在通用服务器平均总功耗与平均切片总时延上的

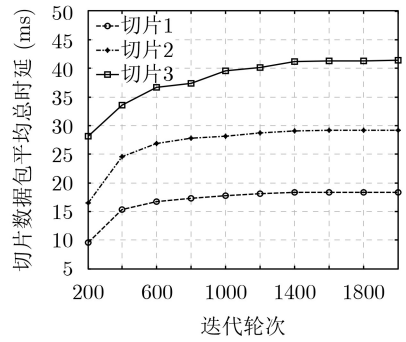

图 3 各切片数据包平均总时延

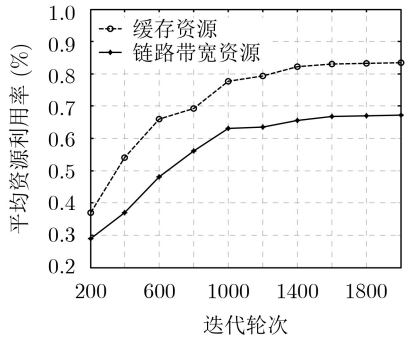

图 4 缓存资源和链路带宽资源平均利用率

性能对比。在此实验中，固定切片1与切片2数据包的平均时延约束分别为20 ms, 30 ms，设置切片3的平均时延约束分别为[40, 45, 50, 55, 60]ms。

从图5与图6中可以看出，随着切片3的平均时延约束增大，不同算法下的通用服务器平均总功耗和平均切片总时延分别减小和增大，但随着平均时延约束达到一定程度之后，通用服务器平均总功耗和平均切片总时延变化速率趋于平缓，这是由于通用服务器平均总功耗优化是以牺牲时延、缓存及带宽资源为代价的，随着切片3平均时延约束的增大，系统会利用VNF共享的机制尽可能减少通用服务器的使用，但由于缓存及带宽资源的约束，通用服务器关闭的数量将达到一个最大值。此时切片3平均时延约束的增大，平均总功耗和平均切片总时延基本保持不变。如图5、图6所示，当切片3的平均时延约束处于较低水平时，基于阈值的VNF迁移整合算法(OVMP)保证了每台开启的服务器资源利用率大于阈值，尽量减少服务器的使用，因此该算法具有最低的平均总功耗，同时，该算法在平均切片总时延的性能表现最差。而随着平均时延约束的增大，其它方案通过对VNF的迁移减少了通用服务器的使用，使得平均总功耗逐渐降低并最终趋于平缓。从图5、图6中可以还看出，DQN算法在平均总功耗和平均切片总时延上整体均处于最优水平。此外，由于状态和行动空间较大，容易出现局部收敛，导致Q学习的方案总体性能低于本文的DQN算法。

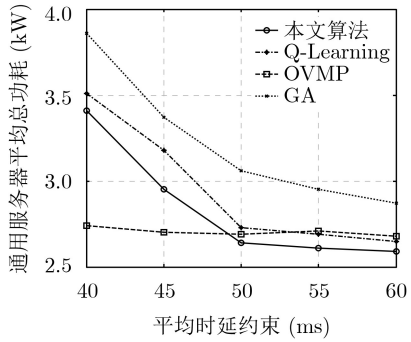

图 5 通用服务器平均总功耗

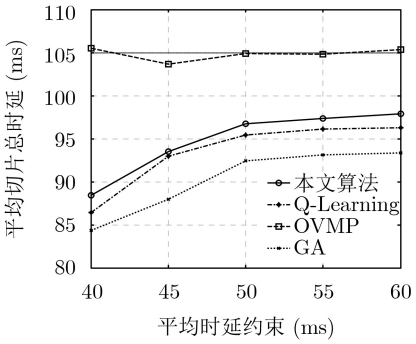

图 6 平均切片总时延

5 结束语

针对5G网络切片架构下业务请求动态变化引起的VNF迁移优化问题，本文基于CMDP建立了一个VNF迁移随机优化模型，进而提出一种基于DQN框架的VNF智能迁移学习算法，本文算法通过CNN来近似行为值函数，从而在每个离散的时隙内根据当前系统状态为每个切片制定合适的VNF编排策略及CPU资源分配方案。仿真结果表明，所提算法在有效地满足各切片QoS需求的同时，降低了基础设施的能耗。由于5G C-RAN架构的提出，后续研究将针对接入网中VNF的迁移以及节点服务性能进行更加精确地建模。

参 考 文 献

[1] GE Xiaohu, TU Song, MAO Guoqiang, *et al*. 5G ultra-dense cellular networks[J]. *IEEE Wireless Communications*, 2016, 23(1): 72–79. doi: 10.1109/mwc.2016.7422408.

[2] SUGISONO K, FUKUOKA A, and YAMAZAKI H. Migration for VNF instances forming service chain[C]. The 7th IEEE International Conference on Cloud Networking, Tokyo, Japan, 2018: 1–3. doi: 10.1109/CloudNet.2018.8549194.

[3] ZHENG Qinghua, LI Rui, LI Xiuqi, *et al*. Virtual machine consolidated placement based on multi-objective biogeography-based optimization[J]. *Future Generation Computer Systems*, 2016, 54: 95–122. doi: 10.1016/j.future.2015.02.010.

[4] ZHANG Xiaoqing, YUE Qiang, and HE Zhongtang. Dynamic Energy-efficient Virtual Machine Placement Optimization for Virtualized Clouds[M]. JIA Limin, LIU Zhigang, QIN Yong, *et al*. Proceedings of the 2013 International Conference on Electrical and Information Technologies for Rail Transportation (EITRT2013)-Volume II. Berlin, Heidelberg: Springer, 2014, 288: 439–448. doi: 10.1007/978-3-642-53751-6_47.

[5] ERAMO V, AMMAR M, and LAVACCA F G. Migration energy aware reconfigurations of virtual network function instances in NFV architectures[J]. *IEEE Access*, 2017, 5: 4927–4938. doi: 10.1109/ACCESS.2017.2685437.

[6] ERAMO V, MIUCCI E, AMMAR M, *et al*. An approach for service function chain routing and virtual function network instance migration in network function virtualization architectures[J]. *IEEE/ACM Transactions on Networking*, 2017, 25(4): 2008–2025. doi: 10.1109/TNET.2017.2668470.

[7] WEN Tao, YU Hongfang, SUN Gang, *et al*. Network function consolidation in service function chaining orchestration[C]. 2016 IEEE International Conference on

Communications, Kuala Lumpur, Malaysia, 2016: 1–6. doi: 10.1109/ICC.2016.7510679.

[8] YANG Jian, ZHANG Shuben, WU Xiaomin, *et al*. Online learning-based server provisioning for electricity cost reduction in data center[J]. *IEEE Transactions on Control Systems Technology*, 2017, 25(3): 1044–1051. doi: 10.1109/TCST.2016.2575801.

[9] CHENG Aolin, LI Jian, YU Yuling, *et al*. Delay-sensitive user scheduling and power control in heterogeneous networks[J]. *IET Networks*, 2015, 4(3): 175–184. doi: 10.1049/iet-net.2014.0026.

[10] LI Rongpeng, ZHAO Zhifeng, CHEN Xianfu, *et al*. TACT: A transfer actor-critic learning framework for energy saving in cellular radio access networks[J]. *IEEE Transactions on Wireless Communications*, 2014, 13(4): 2000–2011. doi: 10.1109/TWC.2014.022014.130840.

[11] WANG Shangxing, LIU Hanpeng, GOMES P H, *et al*. Deep reinforcement learning for dynamic multichannel access in wireless networks[J]. *IEEE Transactions on Cognitive Communications and Networking*, 2018, 4(2): 257–265. doi: 10.1109/TCCN.2018.2809722.

[12] HUANG Xiaohong, YUAN Tingting, QIAO Guanghua, *et al*. Deep reinforcement learning for multimedia traffic control in software defined networking[J]. *IEEE Network*, 2018, 32(6): 35–41. doi: 10.1109/MNET.2018.1800097.

[13] HE Ying, ZHANG Zheng, YU F R, *et al*. Deep-reinforcement-learning-based optimization for cache-enabled opportunistic interference alignment wireless networks[J]. *IEEE Transactions on Vehicular Technology*, 2017, 66(11): 10433–10445. doi: 10.1109/TVT.2017.2751641.

[14] GLOROT X and BENGIO Y. Understanding the difficulty of training deep feedforward neural networks[C]. The International Conference on Artificial Intelligence and Statistics, Sardinia, 2010: 249–256.

[15] PERUMAL V and SUBBIAH S. Power-conservative server consolidation based resource management in cloud[J]. *International Journal of Network Management*, 2014, 24(6): 415–432. doi: 10.1002/nem.1873.

[16] QU Long, ASSI C, SHABAN K, *et al*. Delay-aware scheduling and resource optimization with network function virtualization[J]. *IEEE Transactions on Communications*, 2016, 64(9): 3746–3758. doi: 10.1109/TCOMM.2016.2580150.

唐　伦：男，1973年生，教授，博士生导师，研究方向为新一代无线通信网络、异构蜂窝网络、软件定义无线网络等.

周　钰：男，1993年生，硕士生，研究方向为5G网络切片资源分配和深度学习.

谭　颀：女，1995年生，硕士生，研究方向为5G网络切片、资源分配、随机优化理论.

魏延南：男，1995年生，硕士生，研究方向为5G网络切片、虚拟资源分配，可靠性.

陈前斌：男，1967年生，教授，博士生导师，研究方向为个人通信、多媒体信息处理与传输、下一代移动通信网络.