

# 基于深度信念网络资源需求预测的虚拟网络功能动态迁移算法

唐伦\* 赵培培 赵国繁 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网络场景下缺乏对资源需求的有效预测而导致的虚拟网络功能(VNF)实时性迁移问题, 该文提出一种基于深度信念网络资源需求预测的VNF动态迁移算法。该算法首先建立综合带宽开销和迁移代价的系统总开销模型, 然后设计基于在线学习的深度信念网络预测算法预测未来时刻的资源需求情况, 在此基础上采用自适应学习率并引入多任务学习模式优化预测模型, 最后根据预测结果以及对网络拓扑和资源的感知, 以尽可能地减少系统开销为目标, 通过基于择优选择的贪婪算法将VNF迁移到满足资源阈值约束的底层节点上, 并提出基于禁忌搜索的迁移机制进一步优化迁移策略。仿真表明, 该预测模型能够获得很好的预测效果, 自适应学习率加快了训练网络的收敛速度, 与迁移算法结合在一起的方式有效地降低了迁移过程中的系统开销和服务级别协议(SLA)违例次数, 提高了网络服务的性能。

**关键词:** 虚拟网络功能; 预测; 迁移; 深度学习

中图分类号: TN929.5

文献标识码: A

文章编号: 1009-5896(2019)06-1397-08

DOI: 10.11999/JEIT180666

## Virtual Network Function Migration Algorithm Based on Deep Belief Network Prediction of Resource Requirements

TANG Lun ZHAO Peipei ZHAO Guofan CHEN Qianbin

(School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

(Chongqing Key Laboratory of Mobile Communications Technology, Chongqing 400065, China)

**Abstract:** To solve the problem of real-time migration of Virtual Network Function (VNF) caused by lacking effective prediction in 5G network, a VNF migration algorithm based on deep belief network prediction of resource requirements is proposed. The algorithm builds firstly a system cost evaluation model integrating bandwidth cost and migration cost, and then designs a deep belief network prediction algorithm based on online learning which adopts adaptive learning rate and introduces multi-task learning mode to predict future resource requirements. Finally, based on the prediction result as well as the perception of network topology and resources, the VNFs are migrated to the physical nodes that meet the resource threshold constraints through greedy selection algorithm with the goal to optimize system cost, and then a migration mechanism based on tabu search is proposed to further optimize the migration strategy. The simulation results show that the prediction model can obtain good prediction results and adaptive learning rate accelerates the convergence speed of the training network. Moreover, the combination with the migration algorithm reduces effectively system cost and the number of Service Level Agreements (SLA) violations during the migration process, and improves the performance of network services.

**Key words:** Virtual Network Function (VNF); Prediction; Migration; Deep learning

收稿日期: 2018-07-05; 改回日期: 2019-01-28; 网络出版: 2019-02-19

\*通信作者: 唐伦 tangl@cqupt.edu.cn

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

目前移动网络行业正在迅速向5G演进<sup>[1]</sup>。5G网络切片是虚拟网络中灵活配置资源的技术<sup>[2]</sup>, 每个业务请求即服务功能链(Service Function Chaining, SFC)由一些不同的虚拟网络功能(Virtual Network Function, VNF)组成。大多数文献如<sup>[3-6]</sup>

研究SFC的部署问题, 没有考虑到SFC在其生命周期内的资源需求动态变化问题, 当SFC资源需求超过底层节点的负载阈值时, 会造成服务性能下降, 这就需要实时地迁移出过载底层节点的VNF到其他负载较低的节点上来保障SFC的服务级别协议(Service Level Agreement, SLA)。

为了解决迁移的滞后性问题<sup>[7,8]</sup>, 可行的方法是采用主动式的预测机制。文献<sup>[9]</sup>提出一种基于图形神经网络(Graph Neural Network, GNN)的VNF资源需求预测算法, 但它没有考虑如何使用预测结果来优化资源分配的问题。深度信念网络(Deep Belief Network, DBN)<sup>[10]</sup>作为深度学习的经典算法之一, 由受限制的玻尔兹曼机(Restricted Boltzmann Machine, RBM)概率模型堆叠而成, 解决了传统神经网络<sup>[9,11]</sup>需要大量标签数据以及随着层数的增加, 模型会很快陷入局部最优解的问题。本文拟采用深度信念网络预测SFC的资源需求量。

综上, 本文建立综合带宽开销和迁移开销的系统总开销模型, 并提出一种基于在线学习的深度信念网络资源需求预测模型提前报告SFC的资源需求情况, 根据预测结果设计了基于拓扑感知的启发式迁移算法, 并以得到的最优解作为初始解进一步提出了基于禁忌搜索的迁移优化算法提高算法的效率。

## 2 网络场景

本文考虑NFV编排和控制架构, 如图1所示。不同于文献<sup>[12]</sup>假设VNF仅请求处理资源, 本文同时考虑CPU、内存和带宽资源, 设它们各自的资源利用率阈值分别是 $r_C, r_M, r_B$ 。通过隔离, SFC1中的VNF2和VNF3以及SFC2中的VNF3可以部署到同一节点上, 当该节点在资源利用率超出阈值时, 需要迁移出其上的1个甚至更多VNF到其它负载较低的节点上。本文的目标就是制定一种自发的迁移策略, 实时监控VNF(虚拟链路)的资源需求状态,

通过预测及时发现其所部署的底层节点(链路)中的资源热点, 并根据提前制定的迁移策略选择合适的底层节点(链路)进行VNF(虚拟链路)的动态迁移。

## 3 网络模型

### 3.1 底层网络

本文把底层网络形式化为一个无向图 $G^S=(N^S, L^S)$ , 其中 $N^S$ 表示底层节点集合,  $L^S$ 表示所有底层链路的集合。节点资源包括CPU资源和内存资源, 每单位CPU资源代表处理1个数据包所需的资源。每个底层节点 $m \in N^S$ 的节点CPU容量为 $C_m^S$ , 内存资源容量为 $M_m^S$ , 连接节点 $m$ 和 $n$ 的链路 $l_{mn}$ 的带宽为 $B_{mn}^S$ 。 $L_{mn}^S$ 是节点 $m$ 和 $n$ 之间无环路的路径集合。

### 3.2 SFC请求

SFCs链形式化成有向图, 表示为 $G^V=(N^V, L^V)$ ,  $N^V$ 表示所有的VNF集合,  $L^V$ 表示连接VNF的所有虚拟链路的集合。每个SFC由一些有序的VNF组成, SFC集合表示为 $S=\{s_q|q=1, 2, \dots, Q\}$ , 每个 $SFC \in Q$ 由 $VNF \in N_q^V \subseteq N^V$ 和连接相邻两个VNF  $u$ 和 $v$ 的虚拟链路 $l_{uv} \in L_q^V \subseteq L^V$ 组成。SFC中每个VNF  $u$ 的CPU资源需求为 $C_u^V$ , 内存资源需求为 $M_u^V$ , 虚拟链路 $l_{uv}$ 带宽需求为 $B_{uv}^V$ 。定义一个二进制变量 $A_{mn}^{uv}=\{0, 1\}$ 表示虚拟链路 $l_{uv}$ 是否映射到物理链路 $l_{mn} \in L^S$ 上。由于SFC是由多个有序的VNF组成的, 经过每个VNF时数据流可能会被压缩或者扩张, 从而所需的带宽发生变化。所以一个VNF所需资源与其接下来的链路所需资源有如式(1)的关系

$$B_{uv}^V(t) = C_u^V(t) \cdot L_p / t_{\text{proc}} \quad (1)$$

其中,  $L_p$ 表示包长,  $t_{\text{proc}}$ 表示包处理时间。

## 4 优化目标

本文建立SFC迁移的系统开销模型, 并设计以

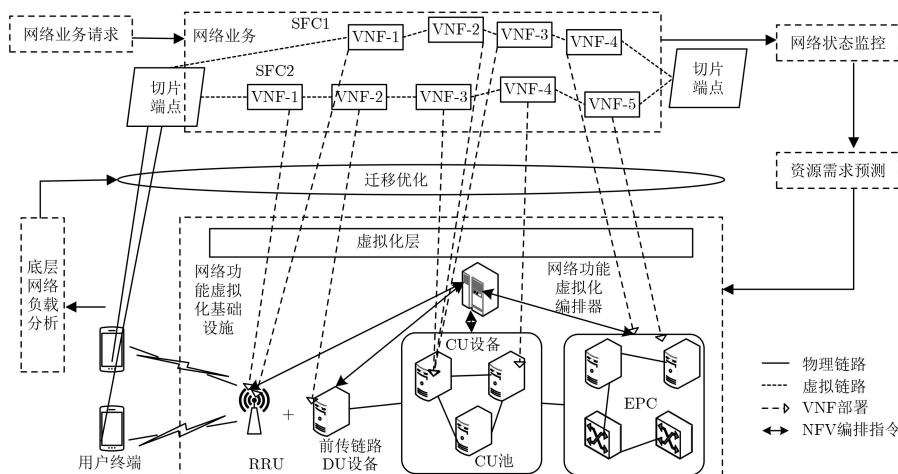


图1 系统场景图

最小化迁移带来的系统开销为目标的迁移算法。定义的系统开销包括两部分：迁移开销和带宽开销。文献[8]指出VNF迁移带来的数据量的传输与VNF迁移带来的能耗开销成正相关关系，而其迁移的数据量主要来源于内存数据的迁移。类似地，本文将迁移VNF的开销近似等价为迁移内存数据所占用网络带宽的时间。这里将VNF  $u$  从底层节点  $n$  迁移到  $m$  的迁移开销定义为

$$C_M^t(u, m) = \sum_{d \in P(m, n)} \frac{M^t(u) \cdot x_n^u(t)}{B^t(d)} \quad (2)$$

其中， $B^t(d)$  表示  $t$  时隙底层节点  $m$  到  $n$  的路径即  $P(m, n)$  上链路  $d$  的剩余可用带宽， $M^t(u)$  表示  $t$  时隙VNF  $u$  的内存资源量。定义VNF  $u$  和  $v$  对底层节点  $m$  和  $n$  产生的带宽开销为

$$C_B^t(u, m, v, n) = B_{uv}^V(t) \cdot x_m^u(t) \cdot x_n^v(t) \cdot \text{hop}^t(m, n) \quad (3)$$

其中， $\text{hop}^t(m, n)$  表示  $t$  时隙底层节点  $m$  到  $n$  的最短距离即所经过的底层链路跳数。定义将  $u$  迁移到  $m$  的带宽开销为

$$C_B^t(u, m) = \sum_{m \in N^S} \sum_{v \in N^V} \sum_{n \in N^S} C_B^t(u, m, v, n) \quad (4)$$

进而将  $u$  迁移到  $m$  的总的系统开销为

$$C_{\text{tot}}^t(u, m) = \alpha \cdot C_M^t(u, m) + \beta \cdot C_B^t(u, m) \quad (5)$$

其中， $\alpha, \beta$  为相应的系数。所以，在  $t$  时隙时虚拟网络功能整体迁移的系统开销为

$$C_{\text{tot}}^t = \sum_{u \in N^V} C_{\text{tot}}^t(u, m) \quad (6)$$

本文的优化目标为最小化迁移带来的系统开销，表示为

$$\begin{aligned} Q1: \quad & \min_{x, A} C_{\text{tot}}^t \\ \text{s.t.} \quad & C1: \sum_{m \in N^S} x_m^u(t) = 1, \quad \forall u \in N^V \\ & C2: \sum_{u \in N^V} x_m^u(t) \cdot C_u^V(t) \leq C_m^S \cdot r_C, \\ & \quad \forall m \in N^S \\ & C3: \sum_{u \in N^V} x_m^u(t) \cdot M_u^V(t) \leq C_m^S \cdot r_M, \\ & \quad \forall m \in N^S \\ & C4: \sum_{l_{uv} \in L^V} (A_{mn}^{uv}(t) + A_{nm}^{uv}(t)) \cdot B_{uv}^V(t) \\ & \quad \leq C_{mn}^S \cdot r_B, \quad \forall l_{mn} \in L^S \\ & C5: \sum_{m \in N^S} A_{nm}^{uv}(t) - \sum_{m \in N^S} A_{mn}^{uv}(t) \\ & \quad = x_n^u(t) - x_n^v(t), \quad \forall n \in N^S, \forall l_{uv} \in L^V \\ & C6: A_{mn}^{uv}(t) = \{0, 1\}, \quad \forall l_{uv} \in L^V, \forall l_{mn} \in L^S \\ & C7: x_m^u(t) = \{0, 1\}, \quad \forall u \in N^V, \forall m \in N^S \end{aligned} \quad (7)$$

本文考虑不可分离流[13]，C1用于保证SFC请求中的每个VNF只能映射到1个底层节点上，C2~C4确保SFC的资源需求不能超过底层节点和链路的资源限制，C5表示SFC中相邻两个VNF映射到底层节点  $m$  和  $n$  时， $mn$  之间必须存在一条连续的路径，C7表示节点映射和链路映射的二进制变量约束。

## 5 基于在线学习的深度信念网络资源需求预测模型

### 5.1 DBN资源需求预测模型

为了优化第4节的目标函数，本文首先使用DBN完成对资源需求的预测。基于在线学习的深度信念网络SFC资源需求预测框架如图2所示，由3部分组成，依次是离线训练、在线学习和在线迁移。

在离线训练阶段，首先对SFC的CPU、内存和带宽资源需求进行特征采集，由于VNF所需的CPU资源和内存资源是相关的，提取CPU资源需求和内存资源需求作为CPU资源需求预测的特征，表示为  $X^{\text{CPU}} = \{C_{\text{VNF}}^V, M_{\text{VNF}}^V\}$ ,  $\forall \text{VNF} \in \text{SFC}$ ，同理，内存资源需求预测的特征表示为  $X^M = \{C_{\text{VNF}}^V, M_{\text{VNF}}^V\}$ ,  $\forall \text{VNF} \in \text{SFC}$ ，带宽的预测可通过公式(1)得到，不进行特征的采集。对于每一条SFC而言，得到的历史观察样本集合表示为  $O = \{\dots O_j \dots\}$ ，其中第  $j$  个样本  $O_j = [X_t, X_{t-1}, \dots, X_{t-d+1}]$ ， $d$  表示样本中时隙的个数，同时也指在线学习中滑动窗口的长度， $X_t = \{X_t^{\text{CPU}}, X_t^M\}$  表示  $t$  时隙时SFC的资源需求特征，对观测数据进行归一化预处理后构造DBN模型对模型参数  $\theta = (\mathbf{w}, \mathbf{a}, \mathbf{b})$  进行正向批训练以提高DBN模型的训练速度，然后进行反向微调过程，构造出最初的预测模型。

在线学习阶段实现对预测模型的实时优化。由于SFC的资源需求监控信息是随时间变化的，一段时间后，最初训练的模型对新样本的预测会出现较大的误差。这里使用离线训练的结果辅助在线学习，采用滑动窗口机制实时更新样本，即每增加一个最新的时隙样本特征，就丢弃一个最旧的时隙样本特征，保持样本大小不变，并使用单样本训练方法重新调整模型参数，然后执行反向微调过程，对模型参数进行优化更新。

最后的在线迁移阶段根据预测的结果判断物理网络中的过载节点，制定迁移策略，进行相应的VNF迁移。迁移完成后，利用监测到的资源需求信息更新样本，从而为下一次的预测以及迁移提供参考。

### 5.2 DBN预测模式

大多数机器学习任务都是单任务学习[10]，各个

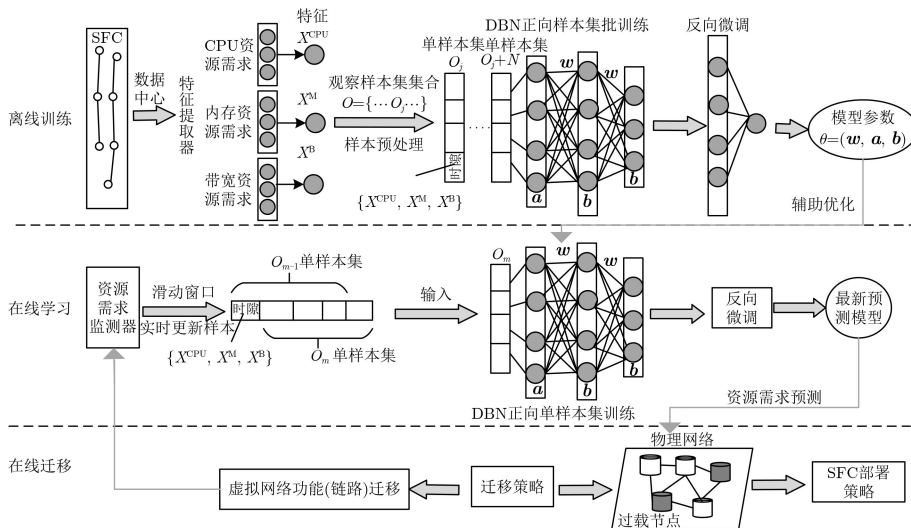


图2 基于在线学习的深度信念网络资源需求预测流程图

任务之间的模型空间(训练模型)是相互独立的,忽略了SFC中VNF之间的丰富关联信息,对模型的预测效果造成影响。本文引入多任务学习MTL<sup>[14]</sup>的方法,如图3所示,它把同一条SFC上的VNF的资源需求同时预测,在共享模型空间中反向联合微调深层产生的特征,可以有更好的预测效果,并改善泛化性能。所以本文采用多任务SFC学习模式建立DBN资源需求预测模型。

5.3 DBN模型参数训练过程

如图4所示,在多任务SFC学习模式的基础上,本文的DBN资源需求预测模型是在文献[10]多层RBM学习模型的基础上堆叠1层多任务回归预测模型,回归层采用sigmoid回归,这样整个结构可以看作一个完整的神经网络结构。本文首先将训练样本 $[X_t, X_{t-1}, \dots, X_{t-d+1}]$ 输入可见层神经元 $v$ 通过非监督贪婪逐层方式进行RBM层参数的训练,然后把RBM层训练得到的特征作为多任务回归层的输入进行预测。由于训练过程中很难得到模型的无偏样本,可以通过 $k$ 步对比散度算法(CD- $k$ )对重构数据进行近似采样,从而得到模型参数 $\theta = (w, a, b)$

的更新公式。一旦在正向无监督阶段的最佳模型参数已经确定,一个有监督的微调阶段就被执行,本文使用BP算法<sup>[11]</sup>对层与层之间的权重和偏置值进行调整。这是通过设置标签值来完成的,它表示SFC中每个VNF真实的CPU资源需求和内存资源需求。最后得到预测输出 $Y_{t+1}$ ,即SFC中每个VNF的CPU资源需求 $\hat{C}_u^V(t+1)$ 和内存资源需求 $\hat{M}_u^V(t+1)$ 。RBM层预训练阶段的学习率 $e^{[10]}$ 对学习过程影响很大,本文通过观察RBM重构误差曲线来判断当前 $e$ 是否合适,从而使得RBM根据实际训练情况自动调整学习率 $e$ 。在训练过程中,若重构误差减小则增大 $e$ ,即乘上1个大于1的数,反之则减小 $e$ ,即乘上1个小于1的数。

6 基于禁忌搜索的动态迁移优化算法(TADM-DBN)

通过第5节预测得到的VNF和虚拟链路下一时刻所需的资源量,计算物理节点和链路的资源利用率,当资源利用率超过该物理网络阈值时,首先采用基于拓扑感知的动态全局迁移算法将虚拟网络功

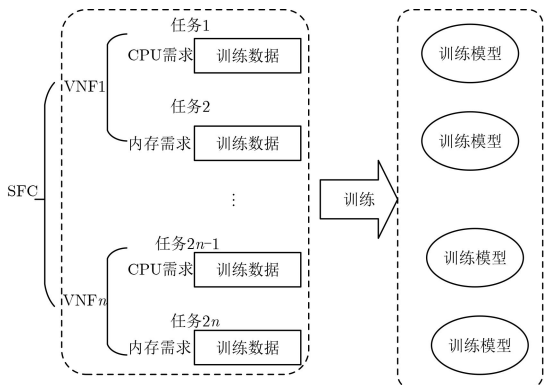


图3 多任务SFC学习模式

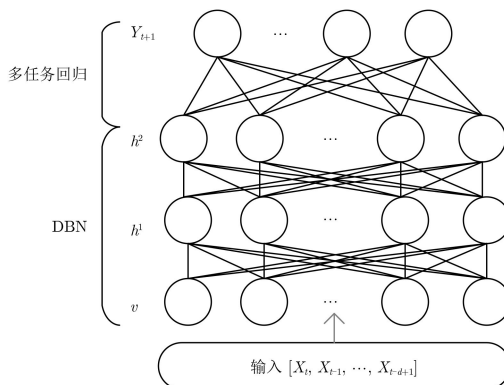


图4 含两个隐藏层的DBN资源需求训练图

能迁移到满足资源阈值约束的底层节点上，然后把得到的解作为初始解通过基于禁忌搜索<sup>[15]</sup>的方法进一步优化迁移策略。

### 6.1 基于拓扑感知的动态全局迁移算法(TPGDM-DBN)

本文提出的基于拓扑感知的动态全局迁移算法如表1的算法1所示，对于预测到的每个过载底层节点，为其上的每个虚拟网络功能贪心地选择目标底层节点。然后提出全局迁移算法如表2的算法2所示，得到最终迁移策略。

算法1中基于拓扑感知的迁移方法主要是指：VNF应迁移到距离其邻居VNF映射节点最近的物理节点上，这里距离是指虚拟链路的重构路径长度，也即跳数。因此拓扑感知的VNF迁移模型可定义为

$$M(u, m) = \left\{ \begin{array}{l} \sum_{v \in \text{neig}(u)} D(m, n), x_m^u = 1, x_n^v = 1 \\ m \in \left\{ m \mid C_m^{S'} > C_u^V \& B^{S'}(p(m, n)) > B_{uv}^V \right\} \end{array} \right\} \quad (8)$$

其中， $D(m, n)$ 表示 $m$ 和 $n$ 之间的满足带宽要求的

表1 基于拓扑感知的动态局部迁移算法(TPLDM-DBN)

算法1 基于拓扑感知的动态局部迁移算法(TPLDM-DBN)	
输入：	过载的物理节点 $S_s$ 以及它上面的VNF集合VNFList
输出：	需要迁移的VNF $m$ 和迁移的目标物理节点 $S_d$
	$C_{\min}[1, 2, \dots,  \text{VNF} ]$ 表示VNFList中每个虚拟机上的VNF $i$ 最小的系统开销
	$D[1, 2, \dots,  \text{VNF} ]$ : $D_i$ 表示给VNF $i$ 带来最小系统开销最小的物理节点
(1)	$C_{\min} \leftarrow \infty$
(2)	for each VNF $i \in \text{VNFList}$ do
(3)	For each $S_j \in S$ do
(4)	If Check_constraints(VNF $i, S_j) = \text{False}$
(5)	Continue;
(6)	endif
(7)	Add $S_j$ to $S_{\text{VNF}_i}$
(8)	Endfor
(9)	对于选出的符合资源约束的物理节点集合 $S_{\text{VNF}_i}$ ，通过拓扑感知计算出VNF $i$ 迁移的最小系统开销 $C_i$ ，并将迁移的目标物理节点用 $D_i$ 表示
(10)	If $C_i < C_{\min}$ and $D_i \neq S_s$ then
(11)	$C_{\min} \leftarrow C_i$
	VNF $m \leftarrow \text{VNF}_i$
	$S_d \leftarrow D_i$
(12)	Endif
(13)	Endfor
(14)	Return (VNF $m, S_d$ )

最短链路跳数。 $C_m^{S'}$ 指节点 $m$ 的可用资源量， $p(m, n)$ 指 $m$ 到 $n$ 的最短路径。 $B^{S'}(p(m, n))$ 表示 $m$ 到 $n$ 的最短路径上的可用带宽量。表中，(10)~(12)行循环的计算出迁移VNFList的每个VNF的最小系统开销， $C_i$ 最小的VNF $i$ 作为最终需要被迁移的VNF $m$ ，并迁移到物理节点 $S_d$ 上。

表2中，算法2的输入为所有过载的物理节点集合及其上的VNF集合，对于每个过载物理节点都执行算法1选择出系统开销最小的VNF进行迁移，直到所有的节点都不会超过资源阈值为止，最后输出相应的解。

表2 基于拓扑感知的动态全局迁移算法(TPGDM-DBN)

算法2 基于拓扑感知的动态全局迁移算法(TPGDM-DBN)	
(1)	输入所有过载的物理节点集合SList以及它们上面的VNF集合VNFList:
(2)	for each $S_i \in \text{SList}$ do
(3)	执行算法1选出迁移的VNF
(4)	while $S_i$ 仍然过载
(5)	继续执行算法2选出系统开销次小的VNF进行迁移
(6)	Endwhile
(7)	Endfor
(8)	输出需要迁移的VNF集合以及相对应的目标迁移物理节点

### 6.2 基于禁忌搜索的动态迁移优化算法(TADM-DBN)

由于上述算法是基于贪婪的启发式算法，得到的解并不是最优解，所以本文采用禁忌搜索算法通过交换过载虚拟网络功能执行迁移策略的顺序进一步优化解的效果。

表3中，算法3通过交换任意VNF迁移顺序产生当前解的邻域候选集合，在候选集合中选取使得系统开销最小的最优解更新初始解，进行新一轮的最优解的搜索。在执行完过载节点的VNF及其相邻的链路迁移后，然后再执行过载底层链路中虚拟链路的迁移，选择具有最小系统开销的虚拟链路进行迁移。

## 7 仿真与性能分析

为了评价模型和算法的有效性，本文在预测性能上与文献[11]中的BP-NN预测算法进行比较，在迁移算法性能上与文献[8]提出的启发式算法进行比较。

### 7.1 仿真设置

本仿真平台在CPU型号为Intel Core i7-4790，内存为8 GB个人电脑上运行，仿真程序在Matlab环境下实现并运行，使用基于GT-ITM工具生成随机的底层网络图。物理节点CPU资源，内存资源和带宽分别均匀地分布在[50, 100] units, [1, 3]

表3 基于禁忌搜索的迁移优化算法(TADM-DBN)

算法3: 基于禁忌搜索的迁移优化算法(TADM-DBN)

- (1) 初始解:  $Z = Z_0, Z^* = Z_0, T = \phi$ , 定义特赦值为  $A(Z^*) = C_{\text{tot}}^t(Z^*)$
- (2) While 不符合终止准则 do
- (3) 产生  $N(Z)$  的一个候选集  $W$ , 在候选集中选取最优解  $X^*$ , 更新  $C_{\text{tot}}^t(X^*)$
- (4) If  $C_{\text{tot}}^t(X) < A(Z^*), X \in T$  且  $C_{\text{tot}}^t(X) < C_{\text{tot}}^t(X^*)$
- (5) 令  $X^* = X$
- (6) 更新  $C_{\text{tot}}^t(X^*)$
- (7) Endif//破禁检查
- (8) If  $C_{\text{tot}}^t(X^*) < C_{\text{tot}}^t(Z^*)$
- (9)  $Z^* = X^*, C_{\text{tot}}^t(Z^*) = C_{\text{tot}}^t(X^*), A(Z^*) = C_{\text{tot}}^t(Z^*)$
- (10) Endif;
- (11) 更新禁忌表  $T, T = T \cup X^*$
- (12) 令  $Z = X^*$
- (13) Endwhile
- (14) Return  $Z^*$

GB和[60, 110] Mbps之间。给定15条SFC请求, 每个SFC请求由1个或多个VNF组成。使用正弦和余弦信号的叠加来模拟SFC资源需求样本数据, 并对其离散化, 为了更加贴近真实环境, 给每个样本加上随机生成的服从泊松分布的数据值, 并定期采集SFC中资源需求数据, 采样周期为1 h, 共采集10000个样本, 前70%作为训练样本进行深度学习训练, 剩下30%的样本作为测试样本进行预测。

## 7.2 资源需求预测

以预测其中1条由3个VNF有序组成的SFC为例进行描述, 本文基于在线学习的DBN资源需求预测模型采用经典的5层模型, 节点数为60-128-32-16-6, RBM的最大迭代次数为300次, 学习率初始值设置为0.0001。本文采用均方根误差(Root Mean Square Error, RMSE)对预测的准确度进行衡量, RMSE定义为

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2} \quad (9)$$

其中,  $Y_t$ 表示标签值,  $\hat{Y}_t$ 表示预测值,  $n$ 为测试样本的个数。RMSE值越小, 预测的准确度越高。选取该SFC中一个VNF的CPU资源需求预测说明本文预测模型的精确度。图5中DBN预测的RMSE值为0.1552, 图6中BP-NN预测的RMSE值为0.2793, 可以看出DBN在预测的拟合精度上要优于BP-NN预测, 这说明本文DBN资源需求预测能够更加有效地从输入数据中提取特征, 提高了资源需求的预测精度, 而BP-NN需要大量标签数据训练

模型参数, 而且随着层数的增加, 模型会很快陷入局部最优解, 从而对预测精度造成影响。此外, 从图7中可以看出对于该SFC中所有VNF的CPU资源需求的预测中多任务学习具有更高的预测精度, 这说明通过把相关任务结合在一起预测的方式能够利用任务间丰富的联系提高预测的整体性能。

本文为了对固定学习率和自适应学习率对RBM收敛速度的影响进行对比, 设定RBM的层数为1层, 自适应学习率的调整为当重构误差减小则把学习率 $\epsilon$ 乘上1.2, 反之把学习率 $\epsilon$ 乘上0.8。如图8所示, 可以看出自适应学习率下的RBM在迭代50次时重构误差就达到了相对平稳, 而固定学习率的RBM迭代200次后才达到相对平稳。因此自适应学习率加快了RBM的收敛速度, 在一定程度上提高了DBN网络的计算效率。

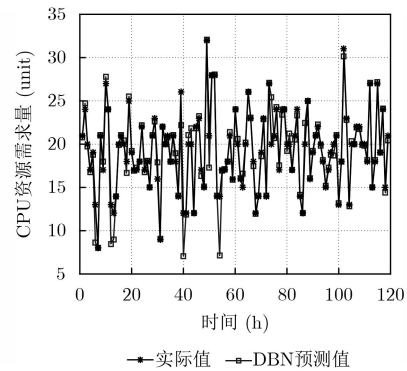


图5 CPU资源需求DBN预测值与真实值趋势对比图

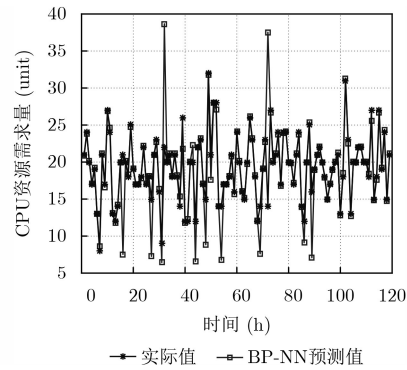


图6 CPU资源需求BP-NN预测值与真实值趋势对比图

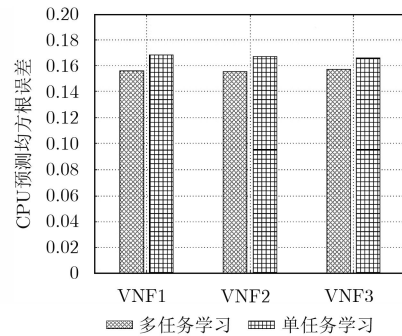


图7 两种学习方式的CPU预测均方根误差对比图

### 7.3 迁移策略

由于本文的迁移开销在物理意义上显示了对服务性能的影响，系统开销越小，服务性能越好。如图9所示，随着时间的推移，系统开销在不断增加，而本文提出的两种算法增加的幅度相对较小。这是因为本文中TPGDM-DBN算法是综合迁移开销和带宽开销进行考虑，使得系统开销得到优化，TADM-DBN采用禁忌搜索方法在TPGDM-DBN的基础上通过选择更优的迁移策略进一步优化了系统开销，而文献[8]中提出的启发式算法只关注带宽开销的优化，没有考虑到迁移内存所带来的系统开销，在性能上相对本文所提算法较差。

为了更加形象的分析阈值的影响，把底层节点的服务性能差异性暂且屏蔽，即通过设置相同的阈值进行仿真。如图10所示，随着阈值的增加，系统开销逐渐减小，这是因为阈值越大，资源需求超过阈值的可能性越小，需要迁移的次数降低，系统开销也会相应减少。而本文两种算法优于文献[8]所提算法。这是因为本文在迁移过程中综合考虑了迁移开销和带宽开销，而文献[8]只优化了带宽开销。

选取3个节点的CPU资源阈值进行分析，其中节点1，节点2和节点3的CPU资源阈值分别为0.65, 0.85和0.55。如图11所示，部署在每个节点中的VNF的CPU资源需求总和都不超过其资源阈值，同理，底层网络的内存资源和带宽资源利用率性能也得到了保障，为了篇幅的需要，这里不进行进一

步的分析。这是因为本文提前通过有效的预测获得了未来时刻的VNF资源需求信息，从而可以在底层节点资源利用率超出阈值前迁移出部分VNF，保证底层节点的资源利用率在可以接受的范围内，从而保障了底层节点的服务性能，满足了SFC业务请求的SLA。图12表示不同算法的SLA违例次数对比图。这里违例次数是指底层网络中资源需求超过阈值情况的发生次数。从图中可以看出本文通过DBN预测显著减少了迁移过程中的违例次数，这是因为通过DBN预测可以提前报告资源需求情况，进而根据预测的需求数据提前做出迁移决策，从而保证SFC服务的性能，提高用户的体验。而基于BP-NN的预测精度相对较低，所以违例次数相对本文较高。而采取了预测机制的算法都比不采取预测机制的算法违例次数显著减少，说明了预测对提高用户体验的重要性。

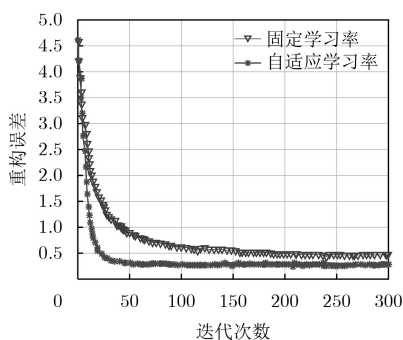


图 8 两种学习率方式的RBM重构误差曲线

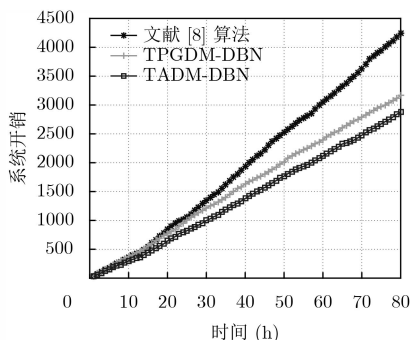


图 9 不同算法系统开销对比图

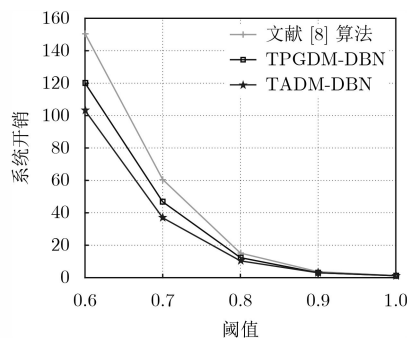


图 10 不同阈值的系统开销对比图

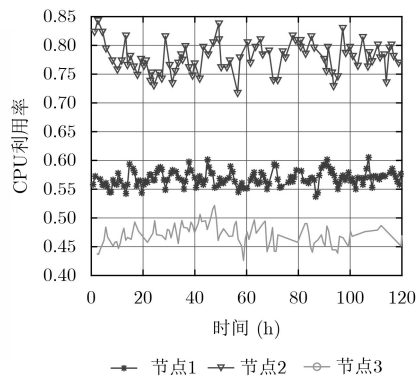


图 11 不同节点的CPU利用率示例图

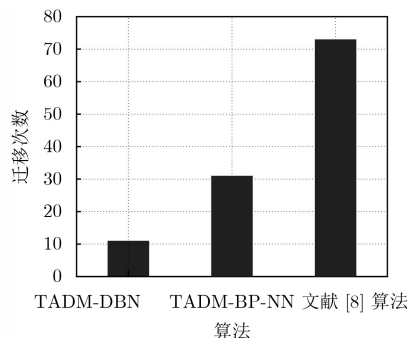


图 12 不同算法的违例次数对比图

## 8 结束语

为了解决SFC服务过程中资源需求的动态性变化带来的服务性能问题, 本文首先建立了综合迁移开销和带宽开销的系统开销模型, 并提出了基于在线学习的深度信念网络资源需求预测算法预测未来的资源需求情况, 本模型引入自适应学习率和多任务学习的模式提升预测效果, 最后利用预测到的结果设计了基于拓扑感知的动态迁移算法, 并以此作为初始解使用禁忌搜索进一步优化迁移决策。仿真结果表明, 本文的预测算法精确度很高, 训练的收敛速度也得到了一定的提升, 此外, 把预测和迁移算法结合在一起的方式有效地降低了系统开销和服务等级协议违例次数, 提高了SFC服务的性能。由于5G C-RAN架构的提出, 协议层的划分不同, 可能会有不同的性能, VNF在接入网的有效部署方案需要提出。所以后续研究中将针对接入网中虚拟网络功能的迁移以及节点服务性能进行更加精确的建模。

## 参考文献

- [1] MAHMOOD N H, LAURIDSEN M, BERARDINELLI G, *et al.* Radio resource management techniques for eMBB and mMTC services in 5G dense small cell scenarios[C]. Proceedings of the 84th Vehicular Technology Conference, Montreal, Canada, 2017: 1–5. doi: [10.1109/VTCFall.2016.7881187](https://doi.org/10.1109/VTCFall.2016.7881187).
  - [2] 唐伦, 张亚, 梁荣, 等. 基于网络切片的网络效用最大化虚拟资源分配算法[J]. 电子与信息学报, 2017, 39(8): 1812–1818. doi: [10.11999/JEIT161322](https://doi.org/10.11999/JEIT161322).  
TANG Lun, ZHANG Ya, LIANG Rong, *et al.* Virtual resource allocation algorithm for network utility maximization based on network slicing[J]. *Journal of Electronics & Information Technology*, 2017, 39(8): 1812–1818. doi: [10.11999/JEIT161322](https://doi.org/10.11999/JEIT161322).
  - [3] RAHMAN M M, DESPINS C, and AFFES S. Design optimization of wireless access virtualization based on cost & QoS trade-off utility maximization[J]. *IEEE Transactions on Wireless Communications*, 2016, 15(9): 6146–6162. doi: [10.1109/TWC.2016.2580505](https://doi.org/10.1109/TWC.2016.2580505).
  - [4] QU Long, ASSI C, SHABAN K, *et al.* A reliability-aware network service chain provisioning with delay guarantees in NFV-enabled enterprise datacenter networks[J]. *IEEE Transactions on Network and Service Management*, 2017, 14(3): 554–568. doi: [10.1109/TNSM.2017.2723090](https://doi.org/10.1109/TNSM.2017.2723090).
  - [5] RIGGIO R, BRADAI A, HARUTYUNYAN D, *et al.* Scheduling wireless virtual networks functions[J]. *IEEE Transactions on Network and Service Management*, 2016, 13(2): 240–252. doi: [10.1109/TNSM.2016.2549563](https://doi.org/10.1109/TNSM.2016.2549563).
  - [6] LUIZELLI M C, BAYS L R, BURIOL L S, *et al.* Piecing together the NFV provisioning puzzle: Efficient placement and chaining of virtual network functions[C]. Proceedings of 2015 IFIP/IEEE International Symposium on Integrated Network Management, Ottawa, Canada, 2015: 98–106. doi: [10.1109/INM.2015.7140281](https://doi.org/10.1109/INM.2015.7140281).
  - [7] XIA Jing, CAI Zhiping, and XU Ming. Optimized virtual network functions migration for NFV[C]. Proceedings of 2016 IEEE International Conference on Parallel and Distributed Systems, Wuhan, China, 2016: 340–346. doi: [10.1109/ICPADS.2016.0053](https://doi.org/10.1109/ICPADS.2016.0053).
  - [8] 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](https://doi.org/10.1109/TNET.2017.2668470).
  - [9] MIJUMBI R, HASIJA S, DAVY S, *et al.* Topology-aware prediction of virtual network function resource requirements[J]. *IEEE Transactions on Network and Service Management*, 2017, 14(1): 106–120. doi: [10.1109/TNSM.2017.2666781](https://doi.org/10.1109/TNSM.2017.2666781).
  - [10] DEUTSCH J and HE D. Using deep learning-based approach to predict remaining useful life of rotating components[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018, 48(1): 11–20. doi: [10.1109/TSMC.2017.2697842](https://doi.org/10.1109/TSMC.2017.2697842).
  - [11] PATI J, KUMAR B, MANJHI D, *et al.* A comparison among ARIMA, BP-NN, and MOGA-NN for software clone evolution prediction[J]. *IEEE Access*, 2017, 5: 11841–11851. doi: [10.1109/ACCESS.2017.2707539](https://doi.org/10.1109/ACCESS.2017.2707539).
  - [12] GIL H J and BOTERO J F. Resource allocation in NFV: A comprehensive survey[J]. *IEEE Transactions on Network and Service Management*, 2016, 13(3): 518–532. doi: [10.1109/TNSM.2016.2598420](https://doi.org/10.1109/TNSM.2016.2598420).
  - [13] MA Wenrui, MEDINA C, and PAN Deng. Traffic-aware placement of NFV middleboxes[C]. Proceedings of 2015 IEEE Global Communications Conference, San Diego, USA, 2015: 1–6. doi: [10.1109/GLOCOM.2015.7417851](https://doi.org/10.1109/GLOCOM.2015.7417851).
  - [14] CARUANA R. Multitask learning[J]. *Machine Learning*, 1997, 28(1): 41–75. doi: [10.1023/a:1007379606734](https://doi.org/10.1023/a:1007379606734).
  - [15] MASHINCHI M H, MASHINCHI M R, and MASHINCHI M. Tabu search solution for fuzzy linear programming[C]. Proceedings of the 7th IEEE/ACIS International Conference on Computer and Information Science, Portland, USA, 2008: 82–87. doi: [10.1109/ICIS.2008.105](https://doi.org/10.1109/ICIS.2008.105).
- 唐伦: 男, 1973年生, 教授, 博士, 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。  
赵培培: 女, 1993年生, 硕士生, 研究方向为5G网络切片映射算法。  
赵国繁: 女, 1993年生, 硕士生, 研究方向为5G网络切片中的资源分配、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、下一代移动通信网络、异构蜂窝网络等。