

# 基于增强学习与 QoE 需求函数的自适应无线网络择取优化机制

高继勋 马小雨

(河南工程学院计算机学院 河南 郑州 451191)

**摘要** 为了使网络择取机制在未知的网络信息条件下仍可适应复杂多变的网络;并可迎合客户动态变化的 QoE 需求,且能对其进行优化,提出了增强学习与 QoE 需求函数相融合的自适应网络择取优化机制。从各类用户出发,划分不同性质的业务,并引入随机离散原则,建立 QoE 需求函数;根据 QoE 反馈信号,兼顾切换决策的主观与客观性,嵌入最小二乘法,构建网络切换决策优化模型,获得决策最优值;计算切换开销与客户需求的概率中间值,并定义了最优网络切换规则,耦合增强学习对网络进行择取/切换;设计增强学习参数更新法,完成会话业务发送。在 MATLAB 仿真平台上对该机制以及其他几种无线网络择取方法进行了测试,实验结果显示:随着网络切换开销权重的不断增大以及转移概率矩阵的实时变化,所提出的机制的性能最好,而且最稳定。

**关键词** 增强学习 QoE 需求函数 决策优化 自适应网络择取机制 开销权重 转移概率矩阵

中图分类号 TP3 文献标识码 A DOI:10.3969/j.issn.1000-386x.2015.07.035

## ADAPTIVE WIRELESS NETWORK SELECTION OPTIMISATION MECHANISM BASED ON REINFORCEMENT LEARNING AND QOE DEMAND FUNCTION

Gao Jixun Ma Xiaoyu

(School of Computer, Henan Institute of Engineering, Zhengzhou 451191, Henan, China)

**Abstract** In order to make the network selection mechanism adapt to the complex and changing network under the condition of unknown network information and to cater for the QoE demand of users dynamic changes, as well as can be optimised, we propose the adaptive network selection optimisation mechanism which integrates the reinforcement learning and QoE demand function. It segments the businesses with different natures from the view of each kind of customer, and introduces random discretion principle to set up QoE demand function; according to the feeding signal from QoE, it takes into account both the subjectivity and objectivity of switching decisions, embeds the least squares method and builds the optimised network switching model as well as gets the optimal decision value; It calculates the compromise probability of switching expense and user needs, and defines the best network switching rules, then couples the reinforcement learning to select and switch the network; Finally, it completes the session service transmission by designing the update method of reinforcement learning. On MATLAB simulation platform this mechanism and several other selection methods for wireless network are tested, and the experimental results show that with the continuously increase of network switching expense weight and the real time change of transition probability matrix, the proposed mechanism has the best performance, and is the most stable one as well.

**Keywords** Reinforcement learning QoE demand function Decision optimisation Adaptive network selection mechanism Expense weight Transition probability matrix

## 0 引言

伴随着无线网络技术的日益发展,世界各国的无线网络通信业务量得到了急剧增加,且用户对无线网络的传输性能提出了更高的要求<sup>[1]</sup>。为了克服这个难题,保证通信网络有一个良好的性能,研究人员开发出了异构无线网络技术<sup>[2]</sup>。异构无线网络是一种包含多种无线接入技术、协议类型、发送方式或者是不同发射功率基站的网络架,能够灵活接入多种不同类型无线网络的能力,具高的无线异构网络资源利用率、低成本、网络容量大以及网络覆盖能力强等优点,得到了诸多学者的广泛研究<sup>[2-4]</sup>。

而异构网络的传输质量由其择取方法控制。故开发一个最

佳的异构网络择取机制以增强其传输性能是非常重要的。如唐良瑞<sup>[5]</sup>等人在综合考虑用户移动性、通话质量等基础上,开发出一种均衡异构无线网络负载的动态接入选择方法,并对方法进行了实验验证,结果显示该机制能够减小候选网络的阻塞率差距,显著提高了网络性能。贾春霞<sup>[6]</sup>等人提出了一种基于模糊层分析法的异构无线网络选择算法,并采用仿真平台验证了其算法的正确性和有效性。Mino G 等人<sup>[7]</sup>从用户速率以及优先级等方面考虑,并引入模糊逻辑,提出了一种新的异构无线网络选择方法,实验数据显示该选择机制可以显著改善异构网络的传输性能。

收稿日期:2013-11-15。高继勋,讲师,主研领域:计算机网络与安全。马小雨,讲师。

尽管以上方法在一定程度上可改善网络传输质量;但由于当前异构无线网络的选择机制都是围绕 QoE 效用函数或者切换函数这两方面,使得这些选择方法只考虑了某个特定的 QoE 需求,无法可满足用户不断变化的 QoE 需求;且难以共享网络资源,在未知的网络状态信息下,无法适应动态变化的网络环境,自适应能力较差。

为了解决上述不足,本文提出了增强学习与 QoE 需求函数相融合的自适应网络择取优化机制。该机制考虑客户和业务的差异化,建立 QoE 需求函数;再由 QoE 反馈信号,嵌入最小二乘法创建网络决策优化模型,根据决策最优值,利用增强学习对网络进行择取与切换,使得本文机制在未知网络信息条件下,可达到自适应动态变化网络目的,并能消除异构网络间的竞争关系,达到资源共享。

## 1 增强学习融合 QoE 需求函数的自适应网络择取优化机制研究

本文择取机制流程如图 1 所示;其工作机制过程如图 2 所示。从图中可以看到,本文机制主要包括三个步骤:(1)建立适用于不同用户和业务的 QoE 需求函数;并初始化增强学习的参数;(2)按照增强学习来择取一个最佳网络;(3)更新增强学习的变量,完成业务传输。

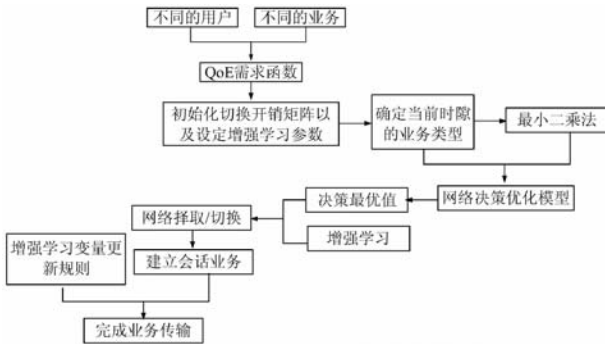


图1 本文自适应无线异构网络的择取机制

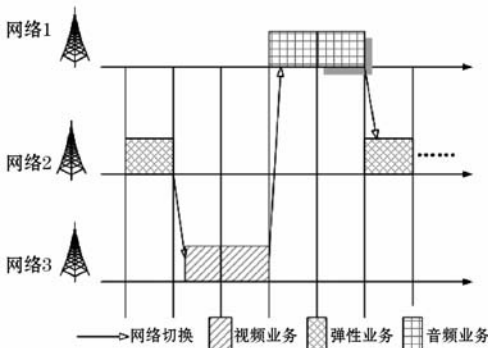


图2 本文择取优化机制工作框架图

本文机制具体设计如下:

### 步骤1 建立 QoE 需求函数<sup>[10]</sup>。

1) 本文针对三种不同的业务:视频;音频;弹性。构建各自对应的 QoE 需求函数;并将利用该函数将网络性能反馈为净回报:

#### (1) 视频业务

将其记为  $B_{Video}$ , 它的 QoE 需求由其最大信噪率  $P_{snr}$  控制。

对应的 QoE 需求函数如下:

$$F_{Video}(P_{snr}) = 4.5 - \frac{3.5}{1 + \exp((b_1(P_{snr} - b_2)))} \quad (1)$$

其中,  $F_{Video}(P_{snr})$  代表该业务的 QoE 需求函数;  $b_1$ 、 $b_2$  均为模型式(1)的两个参数。  $b_1$  用于控制用户体验的扩张速度;  $b_2$  控制  $F_{Video}(P_{snr})$  的正中位置。

#### (2) 音频业务

将其记为  $S_{audio}$ , 主要受数据丢包率  $e$  与时间延迟  $\tau$  的影响。由于异构网络处于动态环境, 考虑随机离散原则; 则二者的计算模型如下:

$$e = e_{network} + (1 - e_{network})e_{playout} \quad (2)$$

$$e_{network} = e_{min} + e_{unit}n_e \quad (3)$$

$$\tau = \tau_{codec} + \tau_{network} + \tau_{playout} \quad (4)$$

$$\tau_{network} = \tau_{min} + \tau_{unit}n_\tau \quad (5)$$

其中,  $e_{network}$  代表数据发送阶段的丢包率;  $e_{min}$  是最低丢包率;  $e_{unit}$  代表  $e_{min}$  的单位;  $n_e$  代表  $n$  个离散状态;  $e_{playout}$  是缓冲该业务引起的丢包率;  $\tau_{codec}$  代表编码器的时间延迟;  $\tau_{network}$  是数据发送引起的时间延迟;  $\tau_{min}$  是最小时间延迟;  $\tau_{unit}$  代表  $\tau_{min}$  的单位;  $n_\tau$  代表  $n$  个延迟值的离散状态; 最小  $\tau_{playout}$  代表缓冲该业务引起的时间延迟。这两个因素对其性能的降低程度记为  $R_e$  与  $R_\tau$ :

$$R_e = \beta_1 + \beta_2 \ln(1 + \beta_3 e) \quad (6)$$

$$R_\tau = 0.024\tau + 0.11(\tau - 177.3)R_{(\tau-177.3)} \quad (7)$$

其中,  $\beta_1$ 、 $\beta_2$ 、 $\beta_3$  均为参数, 其值由所选择的编码器;  $\tau$  用“ms”计量, 177.3 ms 代表  $\tau$  的门限值;  $R_{(x)}$  是指示函数。

再综合  $R_e$  与  $R_\tau$ , 由二者造成的该业务性能降低程度  $I_k$  可用如下方程表示:

$$I_k = 94.2 - R_e - R_\tau \quad (8)$$

根据  $I_k$  值, 推导其对应的 QoE 需求函数  $F_{audio}(I_k)$ :

$$F_{audio}(I_k) = 1 + 0.035I_k + 7 \times 10^{-6}I_k(I_k - 60)(100 - I_k) \quad (9)$$

#### (3) 弹性业务

将它记为  $S_{elastic}$ ;  $S_{elastic}$  对应的 QoE 主要受网络吞吐量  $\alpha$  的影响。考虑不断变化的网络状况, 嵌入随机离散原; 则它的 QoE 需求函数如下:

$$F_{elastic}(\alpha) = b_3 \log(b_4 \alpha) \quad (10)$$

$$\alpha = \alpha_{min} + \alpha_{unit}n_\alpha \quad (11)$$

其中,  $b_3$ 、 $b_4$  均为该函数的参数;  $\alpha_{min}$  代表最低吞吐量;  $\alpha_{unit}$  是  $\alpha_{min}$  的单位;  $n_\alpha$  代表  $n$  个吞吐量的离散状态。计算方程如下:

$$b_4 = \frac{1}{\alpha_1} \sqrt{\frac{\alpha_2}{\alpha_1}} \quad (12)$$

$$b_3 = \frac{1}{\log(b_4 \alpha_1)} \quad (13)$$

其中,  $\alpha_1$  代表最低的预计吞吐量;  $\alpha_2$  是预计吞吐量的峰值。

#### 2) 增强学习参数设置。

用  $S$  这些不同的业务集合:  $S = \{s_{video}, s_{audio}, s_{elastic}\}$ ; 当前网络集合用  $N = \{n_1, n_2, \dots, n_N\}$  表示; 且  $x = [s, n] \in X$  代表网络客户的状态;  $x$  是客户所选网络为  $n$ , 且业务变成  $s$  时的状态;  $X = S \times N$  是客户状态空间;  $\delta \in N$  代表切换后的网络; 增强学习的初始值  $Q(x, n) = 0$ ;  $T$  是网络时隙间距;  $C = \{c_{m,n}\}$ ,  $m, n \in N$  是不同网络间的切换开销矩阵;  $w$  代表开销权重;  $c_{m,n}$  是网络  $m$  变成  $n$  的开销, 且初始化的  $c_{m,n} = 0$ 。

#### 步骤2 建立网络决策优化模型; 根据增强学习择取网络。

① 假设在  $t$  时隙发送的业务种类为  $s(t)$ ;

② 考虑客户在用的网络  $n(t)$ , 可得:

$$x(t) = [s(t), n(t)] \quad (14)$$

③ 确定最优网络切换决策  $\eta(t)$ 。

若切换开销矩阵中各个元素的权重数组为  $W = [w_1, w_2, \dots, w_n]^T$ , 则异构网络中的各网络的最终决策为  $\eta_i$ :

$$\eta_i = \sum_{j=1}^n w_j c_{ij} \quad i = 1, 2, \dots, n \quad (15)$$

其中,  $c_{ij}$  为开销矩阵  $C$  的元素。

本文提出的网络择取机制得到的属性权值体现了客户的主管需求, 但为了兼顾主观和客观, 引入最小二乘法, 则最优网络切换决策模型如下:

$$\begin{cases} \min V(W) = \sum_{i=1}^m \left[ \left( \sum_{j=1}^n w_j c_{ij} - \lambda_i \right)^2 + \left( \sum_{j=1}^n w_j c_{ij} - s_i \right)^2 \right] \\ \text{st. } \sum_{j=1}^n w_j = 1 \\ r_i = \sum_{j=1}^n \sigma_j c_{ij} \quad s_i = \sum_{j=1}^n \phi_j c_{ij} \quad i = 1, 2, \dots, m \\ w_j \geq 0 \quad j = 1, 2, \dots, n \end{cases} \quad (16)$$

其中,  $r_i$  代表开销矩阵的客观决策;  $\sigma_j$  是客观决策特征向量;  $s_i$  代表主观决策;  $\phi_j$  是主观决策特征向量。

根据极限知识, 构造拉格朗日函数:

$$L(W, \varepsilon) = V(W) + 4\varepsilon \left( \sum_{j=1}^n w_j - 1 \right) \quad (17)$$

其中,  $\varepsilon$  代表 Lgrange 因子。令:

$$\frac{\delta L}{\delta w_j} = \frac{\delta V(W)}{\delta w_j} + 4\varepsilon = 0 \quad (18)$$

$$\frac{\delta L}{\delta \varepsilon} = 4 \left( \sum_{j=1}^n w_j - 1 \right) = 0 \quad (19)$$

其中,  $\frac{\delta L}{\delta w_j}$  代表对权重向量元素的导数。

通过式(14)和式(15), 获得  $(n+1)$  维方程组。通过求解可得权重  $W$ :

$$W = J_{nn}^{-1} \left[ K_{n1} + \frac{1 - e_{n1} J_{nn}^{-1} K_{n1}}{e_{n1} J_{nn}^{-1} K_{n1}} \right] \quad (20)$$

$$e_{n1} = [1, 1, 1, \dots, 1]^T \quad (21)$$

$$J_{nn} = \left[ \sum_{k=1}^m c_{ky} c_{kz} \right] \quad y, z = 1, 2, \dots, m \quad (22)$$

$$K_{n1} = \left[ \sum_{i=1}^m \frac{r_i + s_i}{2} c_{i1}, \sum_{i=1}^m \frac{r_i + s_i}{2} c_{i2}, \dots, \sum_{i=1}^m \frac{r_i + s_i}{2} c_{in} \right]^T \quad (23)$$

由上述公式可知其最终决策为  $\eta_i$ ; 再从大到小重排  $\eta_i$ , 确定其对应的最佳网络。

④ 由最优网络切换规则, 耦合增强学习来择取。

(a) 以概率  $\rho(t)$  随机选择一个网络进行连接; (b) 再以概率  $(1-\rho)$  选择网络  $\mu$  接入:

$$\mu = \arg \max_{\eta \in N} Q(x(t), \eta(t)) \quad (24)$$

$$\rho(t) = \frac{3}{5 + \ln t} \quad t \geq 1 \quad (25)$$

**步骤3 增强学习的参数更新**<sup>[10]</sup>

① 完成步骤2后, 开始发送相应的业务。

② 在结束时隙  $t$  的传输后, 依据其传输的业务类别, 得到其对应的 QoE 回报  $F_{s(t)}$ ; 再计算客户的净回报  $z(x(t), \eta(t))$ :

$$z(x(t), \eta(t)) = F_{s(t)} - w c_{n(t), \eta(t)} \quad (26)$$

其中,  $w \in [0, 1]$  代表权重,  $c_{n(t), \eta(t)}$  代表切换开销矩阵的元素。

③ 从  $(t+1)$  时隙起, 业务类别变成  $s(t+1)$ ; 则整个网络状态处于一个新的环境:

$$x(t+1) = [s(t+1), n(t+1)] \quad (27)$$

④ 再根据如下方程更新增强学习法中的  $Q$  值:

$$Q(x(t), \eta(t)) = (1 - \xi_t) Q(x(t), \eta(t)) + \xi_t [z(x(t), \eta(t)) + \zeta \max_{\eta \in N} Q(x(t+1), \eta)] \quad (28)$$

其中,  $\xi_t \in [0, 1]$  代表增强学习法中的控制因子;  $t$  越大, 则其值越小, 且其满足如下关系:

$$\sum_{i=0}^{\infty} \xi_i = \infty \quad \sum_{i=0}^{\infty} (\xi_i)^2 < \infty \quad (29)$$

⑤ 令  $t = t+1$ ; 返回步骤2。

## 2 仿真实验及结果分析

为验证本文自适应网络择取优化机制的优异性和正确性, 将本文机制与其他几种最新的异构网络择取方法在 MATLAB 平台上进行对比测试。全文对照组为: 随机选择方法、文献[8]和文献[9], 为了便于实验, 令随机择取法、文献[8]、文献[9]分别为 A、B、C。将每个时刻的净值回报均值视为衡量标准来评估每种方法的性能。从式(22)可知, 净回报的主要由开销权重  $w$  与概率转移矩阵  $C$  控制。故根据这两个因子来实验。实验环境如图3所示。

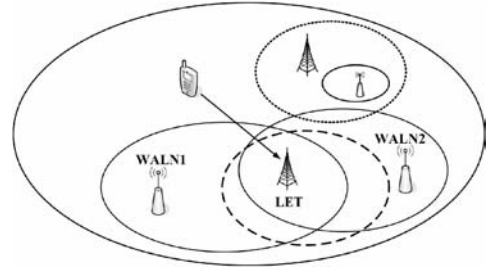


图3 本文实验环境

### 2.1 参数设置

(1) 三个网络的变量值如表1所示。

表1 各变量参数设置

	LET	WLAN1	WLAN2
$e_{\min}$	0.015	0.015	0.03
$e_{\text{unit}}$	0.015	0.015	0.015
$n_e$	3	6	6
$\tau_{\min}$	10 ms	50 ms	60 ms
$\tau_{\text{unit}}$	10 ms	10 ms	10 ms
$n_r$	6	5	6
$\alpha_{\min}$	200 kbps	700 kbps	200 kbps
$\alpha_{\text{unit}}$	50 kbps	65 kbps	50 kbps
$n_\alpha$	8	5	5
$P_{\text{snr}}$	3 dB	6 dB	7 dB

(2) 三类不同业务的 QoE 参数:

① Video 业务:  $b_1 = 1, b_2 = 5$ ; ② audio 业务: 编码器为 G. 711,  $\beta_1 = 0, \beta_2 = 30, \beta_3 = 20, e_{\text{network}} = 0.05, \tau_{\text{network}} = 60$  ms 且  $\tau_{\text{codec}} = 25$  ms; ③ Elastic 业务:  $\alpha_1 = 100$  kbps;  $\alpha_2 = 2$  bMbps; 计算得到  $b_3 = 2.6496, b_4 = 0.0235$ 。

(3) 邻近时刻的转移概率矩阵  $P$ :

$$P = \begin{bmatrix} 0.6 & 0.2 & 0.4 \\ 0.2 & 0.6 & 0.2 \\ 0.3 & 0.4 & 0.5 \end{bmatrix}$$

(4) 不同网络切换开销矩阵  $C$ :

$$C = \begin{bmatrix} & \text{LTE} & \text{WLAN1} & \text{WLAN2} \\ \text{LTE} & 0 & 2 & 2 \\ \text{WLAN1} & 2 & 0 & 1 \\ \text{WLAN2} & 2 & 1 & 0 \end{bmatrix}$$

(5) 增强学习中参数设置: 概率  $\rho(t) = \frac{3}{5 + \ln t}$ ;  $\xi_t =$

$$\frac{4}{3 + 3t^{0.6}}; w = 0.6; T = 30 \text{ ms}; \zeta = 0.2.$$

### 2.2 仿真结果分析

图 4 代表在异构网络满载条件下,不同的择取机制所对应的业务阻塞率均值。从图中可知,在不超过 48 户人数使用时,此时网络有丰富的资源供应给客户,产生阻塞率很低的现象,而这些方法对应的网络性表现非常好的态势;但是当用户超过 50 时,这些方法对应的网络业务阻塞率急剧增加;但是相对于其他几种择取方法而言,本文提出的网络择取优化机制出现的阻塞率较低,只有 0.128%;随机择取方法出现的阻塞率最大,达到 0.252%;而 B、C 分别为 0.193% 和 0.149%。原因是本文方法兼顾了网络的主观和客观性;并依据不同的业务和客户优化了 QoE 需求函数;由最优决策值和增强学习来确定最佳网络。

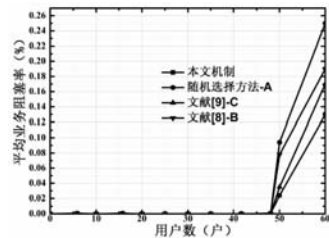


图 4 各网络择取机制对应的业务阻塞率

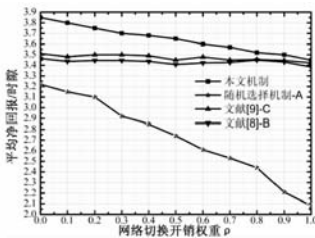


图 5 网络切换开销权重与净回报关系

图 5 是不同机制下的切换开销权重对净回报均值的影响。设置权重的浮动范围为  $[0, 1]$ 。从图中可知,在权重  $w$  不断变化下的条件下,所有的网络择取机制性能都在下降,但本文提出的增强学习与 QoE 需求函数相融合的自适应网络择取优化机制所反馈的净回报均值最大,在这种动态变化的情况,仍表现出优异的网络性能;而尽管 B 和 C 的择取机制所产生的净回报均值没什么变化,但是它们的均值都要小于本文优化机制;而随机择取机制所表现出的性能最差,呈直线下下降趋势,当权重  $w = 1$  时,净回报均值只有 2.08。原因是其他三种网络择取机制无法充分兼容异构网络间的差异性,使其性能受限;而尽管随机择取方法的柔性较好,但该机制不具备学习功能,忽略了网络切换带来的消极影响。文本方法引入了增强学习,优化了网络决策,使得本文机制在不断变化的切换权重下,可以自适应动态变化的环境。本文的网络优化决策如下:

① 当  $0 \leq w \leq 0.4$  时,网络切换决策优化结果为:

	LET	WLAN1	WLAN2
Vedio 业务	WLAN2	WLAN2	WLAN2
audio 业务	LET	LET	LET
elastic 业务	WLAN1	WLAN1	WLAN1

其中,所在的行代表当前网络;列代表当前网络切换为的新业务。若网络是 LET,此时业务是 Vedio,则其最佳网络应切换为 WLAN2;若变为 audio,则最优网络应切换为 LET,即无需切换。

② 当  $0 < w \leq 0.9$  时,网络切换决策优化结果为:

	LET	WLAN1	WLAN2
Vedio 业务	WLAN2	WLAN2	WLAN2
audio 业务	LET	LET	LET
elastic 业务	LET	WLAN1	WLAN1

③ 当  $0.9 < w \leq 1$  时,网络切换决策优化结果为:

	LET	WLAN1	WLAN2
Vedio 业务	WLAN2	WLAN2	WLAN2
audio 业务	LET	LET	LET
elastic 业务	LET	WLAN2	WLAN1

图 6 代表权重  $w = 0.35$  时,在三个不同的业务种类切换矩阵  $P$  条件下,不同的网络择取机制对应的净回报均值。本文采用了如下矩阵  $P_1, P_2, P_3$ :

$$P_1 = \begin{bmatrix} & s_{br} & s_t & s_{el} \\ s_{br} & 0.6 & 0.4 & 0.2 \\ s_t & 0.3 & 0.6 & 0.3 \\ s_{el} & 0.2 & 0.4 & 0.2 \end{bmatrix} \quad P_2 = \begin{bmatrix} & s_{br} & s_t & s_{el} \\ s_{br} & 0.2 & 0.6 & 0.4 \\ s_t & 0.6 & 0.2 & 0.2 \\ s_{el} & 0.5 & 0.4 & 0.3 \end{bmatrix}$$

$$P_3 = \begin{bmatrix} & s_{br} & s_t & s_{el} \\ s_{br} & 0.35 & 0.52 & 0.4 \\ s_t & 0.42 & 0.65 & 0.3 \\ s_{el} & 0.4 & 0.28 & 0.3 \end{bmatrix}$$

从图中可知,随着  $P$  的不断变化,尽管 B、C 以及随机择取法 A 反馈得到的净回报均值只有微小变化;而本文择取机制却呈现降低态势,但是所得净回报均值仍然要大于另外三种方法。这表明在网络动态变化情况下,本文异构网络择取优化机制的性能是较稳定的。

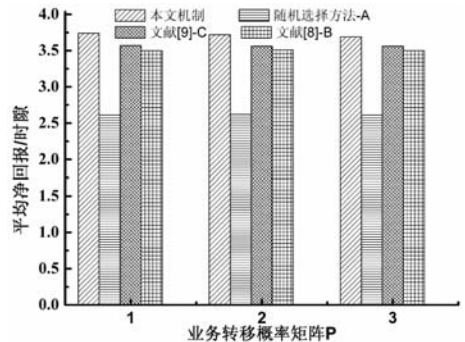


图 6 矩阵  $P$  与净回报均值的关系

### 3 结语

当当前的无线异构网络择取机制都是基于已知先验网络信息,无法在未知条件下适应复杂多变的网络环境;且只考虑了某个特定的 QoE 需求,无法满足客户动态变化的 QoE 需求,故提出了增强学习与 QoE 需求函数相融合的自适应网络择取优化机制。根据不同客户和业务,建立 QoE 需求函数,达到了优化 QoE 目的;构建网络切换决策优化模型,确定决策最优值;计算切换开销与用户需求的折中概率,采用增强学习对网络进行择

取/切换,使得本文机制具有很强的柔性,可在未知信息条件下较好地适应动态变化的网络。仿真结果表明了本文机制在网络切换开销权重以及转移概率矩阵的实时变化状况下,其性能最好,而且最稳定。

### 参 考 文 献

[ 1 ] 许浩,许卫山,高东怀.异构网络安全设备统一管理平台的构建[J]. 科学技术与工程,2012,12(1):1834-1839.

[ 2 ] Charles Sarraf, Firas Ousta, Nidal Kamel. Mapping of QoS between UMTS and WiMAX in Tight Coupling Heterogeneous Wireless Network [J]. International Journal of Soft Computing and Software Engineering,2012,2(3):1-14.

[ 3 ] Jia Chunxia,Liu Jianhui,Liu Shuang. Heterogeneous Wireless Network Selection Based on FAHP [J]. Journal of Xihua University: Natural Science Edition, 2012, 25(1): 156-162.

[ 4 ] Ioannis Chamodrakas, Drakoulis Martakos. A utility-based fuzzy TOPSIS method for energy efficient network selection in heterogeneous wireless networks [J]. Applied Soft Computing, 2012, 12(7): 1929-1938.

[ 5 ] 唐良瑞,杨箏,赵琳.有效均衡异构无线网络负载的接入选择[J].应用科学学报,2013,31(5):441-447.

[ 6 ] 贾春霞,刘建辉,刘爽.基于FAHP的异构无线网络选择[J].西华大学学报:自然科学版,2012(1):41-44.

[ 7 ] Mino G, Ikebata T, HafaF X. The comparison of fuzzy-based intelligent CAC systems and performance evaluation for heterogeneous wireless networks [C]//2010 International Conference on Intelligent Networking and Collaborative Systems, 2010: 177-184.

[ 8 ] 石文孝,李娜,王继红.改进的异构无线网络接入选择方法[J].应用科技学报,2013,31(1):21-26.

[ 9 ] 盛洁,祁兵,杨箏,等.基于粗糙集理论和层次分析方法的异构无线网络接入选择算法[J].计算机应用与软件,2013,30(2):133-136.

[ 10 ] 吴启辉,王金龙,杜智勇,等.异构无线网络中优化用户 QoE 的动态网络选择方法:中国,CN103327556A[P].2013-09-25.

(上接第 92 页)

### 3.2 页面正文提取及评价

获取 U 值后,利用公共子序列进行优化,提取所有行号 j 对应的文本 H[j]. Value(j >= Span[0] 且 j < Span[n]),即为正文内容。本文采用准确率和召回率进行评价。原页面正文的行数为 H[Web-Text]. Count,返回的总行数为 H[Return-Text]. Count,返回有效行数为 H[Valid-Text]. Count。

$$\text{准确率} = \frac{H[\text{Valid-Text}]. \text{Count}}{H[\text{Return-Text}]. \text{Count}} \times 100\%$$

$$\text{召回率} = \frac{H[\text{Valid-Text}]. \text{Count}}{H[\text{Web-Text}]. \text{Count}} \times 100\%$$

实验结果如表 3 所示。

表 3 平均准确率、召回率统计

平均准确率	平均召回率	页面总数(个)
96.35%	93.76%	500

### 3.3 处理效率统计

在正文提取程序的起始处和结束处设置计时标志,记录运行时间。记录结果如表 4 所示。

表 4 处理时间记录

页面总数(个)	处理总时间(秒)	平均处理时间(秒/个)
500	18.975	0.038

### 3.4 实验所用软件截图



图 8 软件截图

## 4 结 语

针对传统页面正文提取方法提取复杂页面正文存在不足的这一现状,通过跟踪主流技术,提出了一种简单的基于统计的复杂页面正文提取方法。页面正文提取的准确度直接影响后续工作。本文所述的方法不依赖页面源码结构,不需要进行大量反复的统计工作且正文提取结果精确高效。

### 参 考 文 献

[ 1 ] 王琦,唐世渭,杨冬青,等.基于DOM的网页主题信息自动提取[J].计算机研究与发展,2004,41(10):1786-1792.

[ 2 ] 胡金栋.网页正文提取及去重技术研究[D].浙江:浙江大学计算机学院,2011.

[ 3 ] 罗刚.使用C#开发搜索引擎[M].北京:清华大学出版社,2012.

[ 4 ] 李效东,顾毓清.基于DOM的Web信息提取[J].计算机学报,2002,25(5):526-533.

[ 5 ] 崔继馨,张鹏,杨文柱.基于DOM的Web信息抽取[J].河北农业大学学报,2005,28(3):90-93.

[ 6 ] 孙晓伟,毕晓清,吴洪越,等.基于DOM和VBA的Word文档数据交互式抽取[J].计算机应用与软件,2013,30(2):113-115.

[ 7 ] 李朝,彭宏,叶苏南,等.基于DOM树的可适应性Web信息抽取[J].计算机科学,2009,36(7):202-203,210.

[ 8 ] 刘军,张净.基于DOM的网页主题信息的抽取[J].计算机应用与软件,2010,27(5):188-190.

[ 9 ] 杨俊,李志蜀.基于DOM的WEB主题信息抽取[J].四川大学学报:自然科学版,2008,45(5):1077-1080.

[ 10 ] 于满泉,陈铁睿,许洪波.基于分块的网页信息解析器的研究与设计[J].计算机应用,2005,25(4):974-976.

[ 11 ] 李蕾,王劲林,白鹤,等.基于FFT的网页正文提取算法研究与实现[J].计算机工程与应用,2007,43(30):148-151.

[ 12 ] 孙皓,董守斌.基于标签密度的自适应正文提取方法[J].郑州大学学报,2009,41(1):44-47.

[ 13 ] 朱泽德,李森,张健,等.基于文本密度模型的Web正文抽取[J].模式识别与人工智能,2013,26(7):667-672.

[ 14 ] 王利,刘宗田,王西华,等.基于内容相似度的网页正文提取[J].计算机工程,2010,36(6):102-104.