

网络虚拟化环境下的业务聚类方法

贾伟, 夏靖波

(空军工程大学信息与导航学院, 西安 710077)

摘要: 为了满足未来丰富多样的网络业务需求, 研究者们提出了许多解决方案, 这些方案的本质都是基于网络虚拟化技术的。本文在网络虚拟化环境下展开业务聚类研究, 并对业务聚类的需求和聚类算法进行了分析, 提出了基于近邻传播聚类算法(AP)的网络业务聚类方法, 针对算法存在的不足, 提出了优化的AP聚类算法(OAP)。实验结果表明采用该方法在业务聚类过程中获得了较好的性能。

关键词: 网络虚拟化; 网络业务; 聚类; 近邻传播

中图分类号: TP393 **文献标志码:** A **文章编号:** 1671-637X(2015)08-0054-06

Methods for Traffic Clustering in Network Virtualization Environment

JIA Wei, XIA Jing-bo

(Information and Navigation College, Air Force Engineering University, Xi'an 710077, China)

Abstract: In order to meet the demand of abundant network traffic in the future, many solutions are proposed by researchers, most of which were based on network virtualization technology. We studied the traffic clustering under network virtualization environment, made analysis to the demand of traffic clustering and some clustering algorithms, and proposed a traffic clustering method based Affinity Propagation (AP) algorithm. Considering the shortcomings of the AP algorithm in traffic clustering, an Optimized Affinity Propagation (OAP) algorithm was designed. Experimental results prove that the OAP algorithm has fine performance in traffic clustering.

Key words: network virtualization; network service; clustering; affinity propagation

0 引言

随着信息技术的不断发展, 网络服务呈现出新型化、多样化的特点, 这对于网络灵活性、安全性和可控性的需求日益突出。但传统基于IP的网络体系结构依赖于无连接分组的模式来实现信息的互联传输, 使得网络无法满足丰富多样的服务需求, 网络僵化问题日益突出。为了解决这个问题, 大量网络功能的修补、增强不断进行, 但是都没有从根本上解决问题, 因此, 发展新型的网络体系结构成为急需解决的问题。

1 网络虚拟化

为了有效解决上述问题, 网络虚拟化技术应运而生, 它一改传统资源与服务紧密耦合的特点, 将网络

基础设施和服务提供相分离, 从而实现业务需求与网络资源的有效匹配, 并能够满足服务的灵活、安全、可控等需求。目前, 国内外很多研究机构针对网络虚拟化展开了大量的研究工作。国外比较著名的项目有美国的 GENI^[1], 它旨在创建一个全新的、安全的、灵活自适应的互联网络; 欧盟的 FIRE^[2] 采用虚拟化的思想, 建设大规模、动态的、可持续的基础设施平台; 日本的 AKARI^[3] 旨在创建一个可随意接入、功能可扩展的新型网络架构。国内的研究有一体化网络与普适服务体系^[4], 旨在实现网络一体化并为用户提供普适服务; 还有可重构信息通信基础网络体系^[5], 旨在提供可扩展、业务普适、可定制、多样化的网络服务。这些研究都为网络虚拟化技术的实现奠定了坚实的基础, 本文就是在网络虚拟化环境下展开业务聚类研究。

2 业务聚类问题描述

在网络虚拟化环境中服务提供是依托基础设施进行承载的, 虚拟网络映射完成了网络资源的分配, 但是

收稿日期: 2015-04-03

修回日期: 2015-04-16

基金项目: 陕西省自然科学基金(2012JZ8005)

作者简介: 贾伟(1986—), 男, 陕西西安人, 博士生, 研究方向为网络虚拟化。

资源的有效分配是建立在对网络业务需求实时感知的基础上的。通过对网络业务进行感知,从中获取业务的优先级、基本属性(源目的IP地址、端口号等)以及服务质量需求(带宽、时延、丢包率和时延抖动等),最终完成资源的有效匹配。然而,不同网络业务的各类属性及需求也是不相同的,如果为每个业务都进行资源的分配难免造成资源的浪费,因此必须对网络业务进行聚类,将具有相似属性及需求的业务划分为一个业务簇,从而为其分配网络资源。具体过程如图1所示。

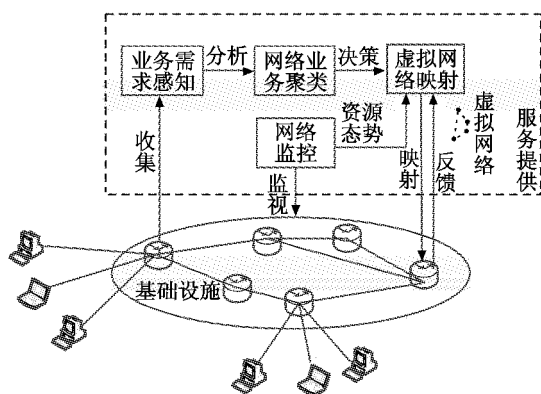


图1 网络虚拟化环境下业务聚类示意图

Fig.1 Traffic clustering under network virtualization environment

网络业务聚类可以分为不同的方式:基于业务的服务质量(如优先级、时延、丢包、抖动、带宽等)、基于业务的通信方式(如单播、组播、多播等)、基于业务的表现形式(如视频、语音、数据等)等。本文主要基于业务的服务质量进行聚类研究。

基于服务质量的虚拟网络业务聚类是将网络业务按照其服务质量需求,通过合适的聚类算法分为多个业务簇,每个业务簇包括一组相似服务质量需求的网络业务,然后构建虚拟网络承载各个业务簇^[6]。其表示为

$$f(A) = C \quad (1)$$

式中: $A = \{a_1, a_2, \dots, a_n\}$,表示一组网络业务的集合,包括 n 个业务,每一个业务都具有不同的服务质量要求; $a_i = \{a_{i,1}, a_{i,2}, \dots, a_{i,q}\}$ 表示第 i 个网络业务的 q 个服务质量需求属性; f 表示聚类方法; $C = \{c_1, c_2, \dots, c_m\}$,表示网络业务通过聚类算法可以分为 m 个业务簇。

3 网络业务聚类算法选择问题分析

“物以类聚,人以群分”,古人很早就提出了聚类的思想。聚类不同于分类,它不需要先验的知识,而是从大量样本中根据一定的相似度计算方法发现样本间潜在的关系,从而确保簇内的样本具有较高的相似性。目前聚类已成为数据挖掘、机器学习、模式识别等领域

的重要内容,国内外很多学者都对聚类算法进行了大量研究,传统的聚类算法有基于划分的聚类^[7](如 k-means)、基于层次的聚类^[8](如 BIRCH)、基于密度的聚类^[9](如 DBSCAN)、基于网格的聚类^[10](如 STING)和基于模型的聚类^[11](如 EM)等。

基于划分的聚类事先将数据集合划分为几个簇,优点是算法简单、速度快,缺点是需要指定簇的个数,不能发现非球形的簇;基于层次的聚类采用凝聚或者分裂的方法构建聚类树,具有能够处理大量数据、收敛速度快的特点,但是同样不能发现非球状的簇;基于数据分布密度进行的聚类,可以发现非球状的簇,但是算法时间和空间复杂度较高;基于网格的聚类采用基于网格的数据结构进行聚类,具有处理速度快的特点,但是降低了聚类的精确度;基于模型的聚类通过构建数据集的概率分布模型来进行聚类,具有输入参数少的特点,但是算法的速率相对较低。

可见,没有一种聚类算法可以保证应用于各类样本空间分析中并保证较好的性能,每一种算法都有自己的特点和应用范围,只有根据具体实际应用去选择合适的聚类算法。

目前对于网络业务的聚类主要集中于网络流量的分析,基于流量来进行业务的分类^[12-13],鲜有基于网络业务服务质量的聚类研究。文献[6]为了解决该问题提出了基于 BIRCH 的业务聚类方法,并针对 BIRCH 不能处理非球状簇的问题结合最近节点距离对算法进行了优化改进。但是改进后的算法仍然存在几个问题:1)算法受数据输入次序的影响较大,这对业务聚类的准确性必然会产生影响;2)算法的每一步操作都与上一步聚类结果相关联,并且操作是不可逆转的,当某一步的聚类结果不合理时,必然会影响后续的聚类结果;3)算法中涉及的参数(B, T, L 等)值难以确定,对于领域的知识依赖性较高。

在网络业务聚类算法的选择过程中,同时需要重点关注以下几个问题。

1) 业务的数量和种类不断增加,因此不能事先确定业务簇的个数^[6],同时要求聚类算法能够处理大量数据。

2) 网络业务对于服务质量的要求越来越高,因此必须保证聚类的准确率与时效性。

3) 网络业务聚类过程中不同属性对于聚类结果的贡献度是不同的,如视频对于丢包率比较敏感、语音对于时延较为敏感。当某一时段内存在大量视频业务时,则需要更加关注丢包率属性。因此应该对属性进行加权处理,从而提高聚类结果准确度。

近邻传播(Affinity Propagation, AP)算法是 FREY

等人 2007 年在 Science 杂志上提出的一种有效的聚类算法^[14],它具有传统聚类算法无法比拟的优点,并广泛应用于人脸识别、文本聚类和图像处理等领域。本文选择 AP 聚类算法用于网络业务聚类,主要基于如下考虑。

1) 算法将所有数据点初始化为候选的聚类中心,从而有效避免了聚类结果受初始聚类中心的影响。

2) 算法具有收敛速度快,适合于处理多类、大规模数据集的特点。

3) 算法基于数据点的相似度来实现聚类,从而避免了 BIRCH 算法由于逐次插入数据带来的输入顺序敏感问题。

4) 算法能够通过参数的调整来动态地控制聚类的粒度,从而使得网络资源的配置更加灵活。

但是传统的 AP 算法同样存在一些缺点,要应用于网络业务聚类还需进行改进和优化。

4 基于优化 AP 算法的网络业务聚类

4.1 基本的 AP 聚类算法

AP 算法将样本空间中所有样本都作为候选的聚类中心,并计算所有样本之间的相似度,通过样本之间信息的传递(归属度和吸引力)迭代获得最优的类代表点,使得数据点到最近类代表点之间相似度和最大。

给定一个数据集 $X = \{x_1, x_2, \dots, x_n\}$, $x_i = \{x_{i1}, x_{i2}, \dots, x_{id}\}$ 代表一个数据点。AP 算法具体过程如下所述。

1) 计算相似度矩阵。这里相似度矩阵基于欧氏距离求得,并取距离的负数,表示距离越近的数据点相似度越大,即

$$s(i, k) = \begin{cases} -\|x_i - x_k\| & i \neq k \\ p(k) & i = k \end{cases} \quad (2)$$

式中, $p(k)$ 表示数据点作为类代表点的偏向程度,值越大表示该点作为类代表点的可能性就越大。由于 AP 算法中数据集的所有点都可能成为潜在的类代表点,因此, $s(k, k)$ 都取相同的值,该值越大,簇的个数越多。

2) 信息传递。信息传递主要依靠两个信息量 $r(i, k)$ 和 $a(i, k)$ 的交替更新来完成。 $r(i, k)$ 表示 x_k 适合作为 x_i 的类代表点的程度, $a(i, k)$ 表示 x_i 选择 x_k 作为其类代表点的合适程度。更新过程为

$$r(i, k) = s(i, k) - \max_{k' \neq k} (a(i, k') + s(i, k')) \quad (3)$$

$$a(i, k) = \begin{cases} \min\{0, r(k, k) + \sum_{i' \neq \{i, k\}} \max\{0, r(i', k)\}\} & i \neq k \\ a(k, k) = \sum_{i' \neq k} \max\{0, r(i', k)\} & i = k \end{cases} \quad (4)$$

为了避免震荡的发生,算法在信息更新过程中引入阻尼因子 λ , $\lambda \in [0, 1)$, 更新迭代过程为

$$\begin{cases} r(i, k)^{(t+1)} = (1 - \lambda) \cdot r(i, k)^{(t)} + \lambda \cdot r(i, k)^t \\ a(i, k)^{(t+1)} = (1 - \lambda) \cdot a(i, k)^{(t)} + \lambda \cdot a(i, k)^t \end{cases} \quad (5)$$

3) 确定类代表点。 x_k 要作为 x_i 的类代表点, k 要满足

$$\operatorname{argmax}(a(i, k) + r(i, k)) \quad (6)$$

4) 终止迭代。当超过规定的最大迭代数或者经过多次迭代类代表点未发生改变则终止迭代。

4.2 AP 算法的缺陷分析

AP 算法具有不用事先指定簇个数、处理速度快等特点,但要应用于网络业务聚类还存在一些缺点。具体如下所述。

1) 算法中数据点相似度是基于欧式距离进行确定的,数据点间距离远的相似度低,距离近的相似度高,这对于处理呈现超球状分布的数据集具有较好的结果,而处理特殊形状和多尺度的数据集则较为困难。

如图 2 所示,在图中双螺旋形状数据集中,根据相似度计算 $s(a, b) > s(a, c)$, 数据点 a 与数据点 b 的相似度要大于与数据点 c 的相似度,即数据点 a 和数据点 b 划为一类的概率较高,但实际情况中数据点 a 和数据点 c 是属于同一类的。因此,为了使算法能够较好地处理具有特殊形状的样本,必须对其相似度矩阵进行改进。

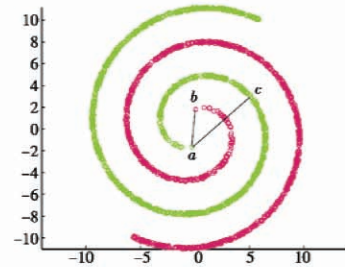


图 2 AP 算法难以处理特殊形状数据集的分簇

Fig.2 Special datasets hard to deal with

2) 算法默认所有属性对于聚类结果的贡献度是一定的。

3) 算法在信息更新过程中为了避免震荡的发生,引入了阻尼因子 λ 来改变算法在不同阶段的搜索能力。但是 AP 算法通常设 λ 为固定值,使得算法无法动态调节不同阶段的搜索性能,从而影响收敛性能。

4.3 优化的 AP 算法

针对网络业务聚类对于聚类算法要求以及 AP 算法存在的缺陷,本文对其进行优化改进,提出了优化的 AP 算法(Optimized Affinity Propagation, OAP)。

4.3.1 相似度计算方法的改进

高斯核函数能够将样本空间非线性映射到高维空间中从而实现样本特征的线性可分,有利于发现非凸的样本空间。但是基于高斯核函数的相似度计算由于

仍然基于核空间的欧氏距离度量方法,并且对尺度参数比较敏感,因此对于处理数据分布密度不均匀的问题仍然存在困难。为此,本文引入文献[15]提出的相似度计算方法,即

$$s(x_i, x_j) = \exp\left(\frac{-d^2(x_i, x_j)}{2\bar{\sigma}^2}\left(1 + \frac{|\sigma_i - \sigma_j|}{\sigma_{\max}}\right)\right) \quad (7)$$

式中: σ_i 表示数据点 x_i 到所处邻域的密度; $\bar{\sigma} = \frac{1}{n} \sum_{i=1}^n \sigma_i$ 表示数据点 x_i 到其 k (这里 k 取7)个最近邻的平均距离; σ_{\max} 取 $|\sigma_i - \sigma_j|$ 中的最大值,这样通过 $1 + \frac{|\sigma_i - \sigma_j|}{\sigma_{\max}}$ 的调整,密度相差越大,相似度越小。

4.3.2 属性赋权

不同网络业务中不同属性对于聚类结果的重要度存在差别,本文引入权重因子 ω_m ^[16],表示数据集中第 m 个属性对于聚类的重要度,其表示为

$$v_m = \left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_m)\right)^{1/2} / |\bar{X}_m| \quad (8)$$

$$\omega_m = v_m / \sum_{m=1}^d v_m \quad (9)$$

式中, $\bar{X}_m = \frac{1}{n} \sum_{i=1}^n x_{im}$, $m \in (1, 2, \dots, d)$,表示数据集中第 m 个属性取值的平均值。

基于上述分析新的相似度计算方法为

$$s(i, k) = \begin{cases} \exp\left(\frac{-\sum_{m=1}^d \omega_m \cdot (x_{im} - x_{km})^2}{2\sigma'^2}\left(1 + \frac{|\sigma_i - \sigma_j|}{\sigma_{\max}}\right)\right) & i \neq k \\ p(k) & i = k \end{cases} \quad (10)$$

4.3.3 自适应阻尼因子策略

在信息更新过程中为了避免发生震荡,引入了阻尼因子 λ ,但是 λ 值设置合适与否会直接影响算法的收敛性能。AP算法中 λ 值为固定常数。本文采用文献[17]提出的两阶段非线性凹函数策略,可以有效地避免震荡,加快收敛。具体算式为

$$\lambda = \begin{cases} -(\lambda_{\text{initial}} - \lambda_{\text{final}})(t/t_{\text{break}})^2 + \lambda_{\text{initial}} & 0 \leq t \leq t_{\text{break}} \\ \lambda_{\text{final}} & t_{\text{break}} < t \leq t_{\text{max}} \end{cases} \quad (11)$$

式中: λ_{initial} 为阻尼因子的初始值; λ_{final} 为阻尼因子的最终值; t_{break} 为两阶段的分界点; t_{max} 为最大迭代次数。

4.3.4 算法的工作流程

算法的具体工作流程如下所述。

输入:数据集 X ,相关参数 λ_{initial} , λ_{final} , t_{max} , t_{break} 和 C_{convits} 。

输出:聚类结果 $C = \{c_1, c_2, \dots, c_m\}$ 。

1) 初始化 $r(i, k)$ 和 $a(i, k)$ 为0,设置算法的相关参数;

2) 根据式(10)计算相似度矩阵;

3) 根据AP算法的原理按照采用自适应阻尼因子策略的式(5)进行迭代更新,确定类代表点;

4) 观察是否满足迭代中止的条件,若满足,输出结果,若不满足,转至步骤3)。

5 实验分析

本节验证OAP算法较传统AP算法在处理不规则形状样本空间以及聚类准确率上的进步,并分别基于具有不规则形状的人工数据集和标准数据集进行了验证,然后模拟真实网络的业务,分析业务对于服务质量的需求。针对这些具体的需求数据集,通过OAP算法进行聚类,验证算法的应用效果。算法中的参数设置为: $p(k)$ 取相似度矩阵的中值, $\lambda_{\text{initial}} = 0.5$, $\lambda_{\text{final}} = 0.9$, $t_{\text{max}} = 2000$, $t_{\text{break}} = C_{\text{convits}} = 100$ 。

5.1 评价标准

为了更好地评价算法的聚类性能,本文选取FMI指标作为评价指标。FMI结合聚类的准确率和查全率来综合评价

$$Prec(t, C_k) = \frac{N_{tk}}{N_k} \quad (12)$$

$$Rec(t, C_k) = \frac{N_{tk}}{N_t} \quad (13)$$

式中: N_k 表示聚类结果中簇 k 的样本数; N_t 表示数据集中类别 t 的样本数; N_{tk} 表示聚类结果中簇 k 包含类别 t 的样本数。那么FMI指标定义为

$$F(t, C_k) = \frac{2Prec(t, C_k) \cdot Rec(t, C_k)}{Prec(t, C_k) + Rec(t, C_k)} \quad (14)$$

整个划分的FMI值为

$$F(C) = \sum_{t \in T} \frac{N_t}{N} \max(F(t, C_k)) \quad (15)$$

FMI取值范围为 $[0, 1]$,值越大,聚类结果越准确。

5.2 人工数据集

实验选择了两种具有不规则形状的数据集DS1和DS2,如图3所示。两个数据集具有不规则的形状,用来比较改进算法对于不规则形状数据簇的处理能力。

图4是优化算法与传统AP算法在两种数据集上的聚类结果。从图中可以看出,传统的AP算法在数据集DS1和DS2上都出现了错误的划分,这主要是由于AP算法基于欧氏距离来计算相似度从而导致错误的划分,而优化的AP算法基本上可以得出正确的聚类结果。由此可见,OAP算法在处理不规则形状数据集时具有明显的优势。



图 3 具有不规则形状的人工数据集
Fig. 3 Artificial datasets with irregular shape

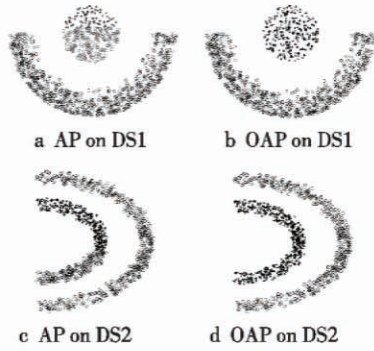


图 4 两种算法在人工数据集上的聚类效果
Fig. 4 Clustering result on artificial dataset of AP and OAP algorithm

5.3 标准数据集

为了验证算法的运行时间和聚类准确率,选取了 3 种 UCI 机器学习数据库标准数据集 Iris, Wine 和 Glass, 各个数据集的特征如表 1 所示。

表 1 数据集特征

Table 1 Feature of dataset

数据集	样本数	维数	类别
Iris	150	4	3
Wine	178	3	3
Glass	214	9	6

在标准数据集运行两种算法,得到如表 2 所示结果。

表 2 两种算法在标准数据集上的运行结果

Table 2 Operation result of AP and OAP algorithm on standard dataset

数据集	算法	运行时间/s	FMI
Iris	AP	0.22	0.884 6
	OAP	0.20	0.905 2
Wine	AP	0.83	0.852 1
	OAP	0.82	0.892 8
Glass	AP	2.31	0.792 1
	OAP	2.39	0.834 6

从表 2 中可以看出,在运行时间上 OAP 和 AP 算法相差不大,在数据集 Glass 上 OAP 算法还高于 AP 算法。这主要是因为 OAP 算法采用了自适应阻尼策略,加快了收敛速度,但是却增加了相似度计算的复杂度;在准确率上,OAP 由于对相似度计算方法进行了优化,从而获得更高的聚类精度。

5.4 网络业务数据集

5.2 节和 5.3 节验证了算法的聚类性能。本节主要针对网络业务数据集进行分析,根据 DiffServ^[18] 定义的业务分类标准生成 2000 种网络业务数据集。数据集中包括时延、抖动和丢包率 3 种服务质量需求。每种网络业务都根据业务分类标准分配相应的类标(控制类为 1,实时类为 2,质量保障类为 3,尽力而为类为 4),用以评估聚类结果的准确率。部分网络业务数据如表 3 所示。

表 3 部分网络业务数据

Table 3 Some data of network traffic

业务类别	丢包率/%	时延/s	抖动/ms
3	0.132 1	1.130 3	175.337 0
2	0.013 5	0.043 7	25.926 8
2	0.002 6	0.054 0	10.550 3
1	0.050 3	0.618 5	92.803 7
1	0.030 1	0.593 0	60.117 8
2	0.016 4	0.068 3	22.424 3
4	0.461 3	16.005 3	126.598 1
3	0.144 4	1.962 3	146.613 7
4	0.974 8	17.681 5	143.992 3
1	0.063 1	0.421 4	90.405 1

在生成的网络业务数据集上分别运行两种算法得到如图 5 所示结果。从图中可以看出,OAP 算法较 AP 算法在网络业务聚类的准确率上有明显的提升,这主要得益于 OAP 算法对 AP 算法采取了一定的优化。相似度计算方法的改进可以使算法能够更准确地处理不规则形状的数据簇,属性赋权使算法能够更好地处理具有不同权重的业务数据集,自适应阻尼策略可以加速算法的收敛速度,从而更好地发现近似最优解。

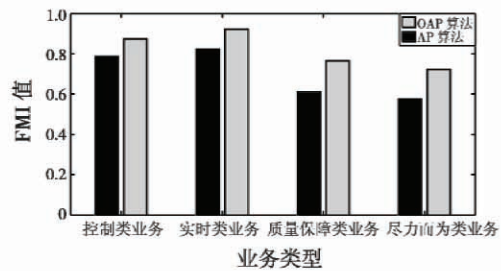


图 5 两种算法在网络业务数据集上的聚类结果

Fig. 5 Clustering result of AP and OAP algorithm on network traffic datasets

6 结束语

本文研究了网络虚拟化环境下基于服务质量的业务聚类问题,针对网络业务聚类过程的需求,提出了基于 AP 算法的网络业务聚类方法,并针对传统 AP 算法在网络业务聚类过程中存在的不足,对 AP 算法进行改进,提出了基于 OAP 算法的网络业务聚类,最后通

过实验验证本文方法获得了较好的聚类结果。下一步工作是在业务聚类的基础上,研究虚拟网络映射问题,实现需求与资源的有效匹配。

参 考 文 献

[1] ELLIOTT C. Exploring network of the future[DB/OL]. [2011-09-07]. http://www.ieice.org/~nv/Chip_Elliott-Spet_10_2011.pdf.

[2] FIRE. Future internet research and experimentation initiative[DB/OL]. [2012-07-10]. <http://cordis.europa.eu/fp7/ict/fire>.

[3] Anon. The overview of AKARI architecture design project [Z]. http://www.nict.go.jp/en/photonic_nw/archi/akari.

[4] ZHANG H K, SU W. Fundamental research on the architecture of new network: universal network and pervasive service[J]. Acta Electronica Sinica, 2007, 35 (4) : 593-598.

[5] LAN J L, CHENG D N, HU Y X. Research on reconfigurable information communication basal network architecture[J]. Journal on Communications, 2014, 35 (1) : 128-139.

[6] 江逸茗, 兰巨龙, 郭通, 等. 一种面向可重构网络的业务聚类方法[J]. 电子与信息学报, 2013, 35 (3) : 703-708. (JIANG Y M, LAN J L, GUO T, et al. An approach of traffic clustering for reconfigurable network[J]. Journal of Electronics & Information Technology, 2013, 35 (3) : 703-708.)

[7] YU S, TRANCHEVENT L C, LIU X H, et al. Optimized data fusion for kernel k-means clustering [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34 (5) : 1031-1039.

[8] ZHANG T, RAMAKRIHNNAN R, LIVNY M. BIRCH: an efficient data clustering method for very large databases [J]. ACM SIGMOD Record, 1996, 25 (2) : 103-114.

[9] ANDRADE G, RAMOS G, MADEIRA D, et al. G-DB-SCAN: a GPU accelerated algorithm for density-based clustering[J]. Procedia Computer Science, 2013, 18 (1) : 369-378.

[10] MUNTZ R R, WANG W, YANG J. STING: a statistical information grid approach to spatial data mining [C]//

The 23rd International Conference on Very Large Data Bases (VLDB), Athens, Greece: Morgan Kaufmann Publishers, 1997: 186-195.

[11] DEMPSTER A P, LAIRD N M, RUBIN D B. Maximum-likelihood from incomplete data via the EM algorithm [J]. Journal of the Royal Statistical Society, 1977, 39 (1) : 1-38.

[12] GOMES J V, INÁCIO P R M, PEREIRA M, et al. Exploring behavioral patterns through entropy in multimedia peer-to-peer traffic [J]. The Computer Journal, 2012, 55 (6) : 740-755.

[13] DAINOTTI A, PESCAPE A, SANSONE C. Issues and future directions in traffic classification [J]. IEEE Network, 2012, 26 (1) : 35-40.

[14] FREY B J, DUECK D. Clustering by passing messages between data points [J]. Science, 2007, 315 (5814) : 972-976.

[15] 王雅琳, 陈斌, 王晓丽, 等. 基于密度调整的改进自适应谱聚类算法 [J]. 控制与决策, 2014, 29 (9) : 1683-1687. (WANG Y L, CHEN B, WANG X L, et al. Improved adaptive spectral clustering algorithm based on density adjustment [J]. Control and Decision, 2014, 29 (9) : 1683-1687.)

[16] 韩旭明, 孙海波, 王丽敏. 基于变异赋权的吸引力传播算法 [J]. 吉林大学学报: 理学版, 2014, 52 (3) : 552-555. (HAN X M, SUN H B, WANG L M. Affinity propagation algorithm based on coefficient of variation weighting [J]. Journal of Jilin University: Science Edition, 2014, 52 (3) : 552-555.)

[17] 王羨慧, 陈俊英, 覃征, 等. 基于动态阻尼因子的动态阻尼仿射传播聚类 [J]. 华中科技大学学报: 自然科学版, 2012, 40 (2) : 116-122. (WANG X H, CHEN J Y, QIN Z, et al. Dynamic damping factor—based dynamic damping affinity propagation clustering [J]. Journal of Huazhong University of Science and Technology: Natural Science Edition, 2012, 40 (2) : 116-122.)

[18] BABIARZ J, CHAN K, BAKER F. Configuration guidelines for DiffServ service classes [DB/OL]. [2015-04-01]. <http://wenku.it168.comld.001568384.shtml>.

