

## 基于用户偏好预测的无人机部署和缓存策略

任佳智<sup>1</sup>, 田辉<sup>1</sup>, 范绍帅<sup>1</sup>, 林远卓<sup>1</sup>, 聂高峰<sup>1</sup>, 李继龙<sup>2</sup>

(1. 北京邮电大学网络与交换技术国家重点实验室, 北京 100876; 2. 国家广播电视总局广播电视科学研究院, 北京 100866)

**摘要:** 针对蜂窝网络中的缓存问题, 考虑用户内容请求的空间异构性及时间波动性, 提出了一种基于单个用户内容偏好预测的蜂窝网中无人机位置部署及缓存内容部署方案。首先基于用户的历史上下文信息, 利用文件相似性及用户相似性来预测每个用户的内容偏好特性, 并使用一种基于线性回归的方法来预测用户未来发起内容请求时的位置和时间; 然后根据预测的地理位置、请求时间和内容偏好, 分别利用基于自组织映射神经网络 (SOM) 的聚类算法和基于凝聚嵌套 (AGNES) 的分簇算法设计无人机的部署位置, 并根据相应的无人机位置设计内容部署方案。仿真结果表明, 所提算法在缓存命中率和时延性能上均优于对比算法。对真实数据集的分析结果表明, 不同的用户特征对内容偏好影响权重不等, 因此需要对不同的用户特征赋予合理的权值。

**关键词:** 边缘缓存; 无人机; 用户偏好预测; 相似性

**中图分类号:** TN929.53

**文献标识码:** A

**doi:** 10.11959/j.issn.1000-436x.2020104

## UAV deployment and caching scheme based on user preference prediction

REN Jiazhi<sup>1</sup>, TIAN Hui<sup>1</sup>, FAN Shaoshuai<sup>1</sup>, LIN Yuanzhuo<sup>1</sup>, NIE Gaofeng<sup>1</sup>, LI Jilong<sup>2</sup>

1. State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

2. Academy of Broadcasting Science, National Radio and Television Administration, Beijing 100866, China

**Abstract:** In order to design an efficient edge caching policy considering spatial heterogeneity and temporal fluctuations of users' content requests, a proactive caching scheme was proposed with UAV's deployment location design based on user preference prediction. Firstly, each user's preference characteristics were predicted based on file similarity and user similarity, and the request time and user location were also predicted when a content request occurs. Thereafter, on the basis of the predicted geographical location, request time and user preference, each UAV's deployment location and the corresponding content placement were determined by virtue of clustering method based on SOM and AGNES. Simulation results show that the proposed scheme outperforms other three comparison schemes in terms of hit ratio and transmission delay. Furthermore, the results also reveal that content preference is correlated with different user features by different weights. Accordingly, different impact weights should be matched with different user features.

**Key words:** edge caching, unmanned aerial vehicle, user preference prediction, similarity

### 1 引言

移动智能终端及相应的各种各样的移动应用的激增使无线数据业务量急剧增长, 并且业务增长模式的驱动力已经由以连接为中心的通信 (例如语

音和短消息) 向以内容为中心的通信 (例如视频流和音乐) 转移。为提升网络容量以满足需求, 各大运营商部署了密集的小基站, 但这对回传链路造成了极大的负担<sup>[1]</sup>。现有的回传网络已无法应对移动业务量的爆炸式增长。第五代移动通信系统采用基

收稿日期: 2019-10-17; 修回日期: 2020-04-14

通信作者: 田辉, tianhui@bupt.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61801044)

**Foundation Item:** The National Natural Science Foundation of China (No.61801044)

于虚拟化的网络架构,使网络资源的使用更加灵活<sup>[2]</sup>。此外,未来网络也正向智能化发展<sup>[3]</sup>。因此,可以在网络边缘(例如无人机、中继或者用户终端)智能地缓存一些流行内容,对流行内容的请求直接由边缘节点提供传输服务,而不需要回传链路的重复传输<sup>[4]</sup>。

通常情况下,用户是移动的,因此需要一个更加灵活的缓存策略。无人机(UAV, unmanned aerial vehicle)作为一种可飞行基站,能够动态地缓存流行内容,追踪无线终端的移动特性,从而有效地为用户提供传输服务。这种方式不仅能够极大地降低传输时延,同时也缓解了高峰时段业务负载对回传链路的负荷压力<sup>[5-7]</sup>。无人机辅助的边缘缓存技术可以将内容直接缓存在无人机中再传输给用户,也可以缓存在用户终端中由无人机进行调度<sup>[8]</sup>。对于前者,内容可在网络闲时缓存在无人机中;对于后者,移动用户缓存被请求的文件,并在地面基站或者无人机的调度下,通过终端到终端(D2D, device-to-device)通信分发内容。这种边缘缓存策略可以在增强用户体验(QoE, quality of experience)的同时降低对回传链路容量的需求。UAV在许多领域具有较广泛的应用,近年来,无线通信领域对UAV的研究也较活跃。例如,第三代合作伙伴计划(3GPP, The 3rd Generation Partnership Project)标准化组织考虑如何在未来的蜂窝网络中支持UAV的无缝集成<sup>[9-15]</sup>。UAV可以通过不同的方式来增强蜂窝通信。一方面,专用UAV可以作为空中无线接入点或者中继节点,进一步提升陆地通信能力,可被称为UAV辅助的蜂窝通信。另一方面,UAV可以作为感知目的,与传统的固定传感节点相比,其优势在于按需部署、较大的服务覆盖面积及灵活的空中网络架构,可被称为蜂窝辅助的UAV感知。因此,有必要在未来的无线蜂窝网络中发展新的通信、信号处理和优化架构来支持高数据速率的UAV通信系统,进而辅助陆地蜂窝通信,并支持超可靠且实时的感知应用<sup>[16-19]</sup>。

文献[20]研究了在云接入网中具有缓存功能的无人机主动部署问题,其目标是 minimized 无人机传输功率。文献[21]提出了一种主动缓存机制来延长无人机的持续使用时间,其中无人机会主动传输文件到选定的地面节点中,这些地面节点通过协作的方式来缓存所有文件。然后,每个地面节点可直接从

本地缓存获得内容或者通过D2D通信从邻近终端获得内容。文献[22]研究了超密集网络中无人机缓存技术中的安全传输问题,其中视频内容被同时缓存在无人机和小基站侧,使用干扰对齐来消除地面用户之间的干扰,并且使用空闲小基站产生干扰信号来破坏潜在的窃听行为。考虑到实际传播群组(例如视距和非视距)的影响,文献[23]设计了一种在多无人机网络中联合无人机缓存内容和服务位置的机制,其目标是达到用户服务概率和传输开销的折中,该机制中最优化问题被建模为一种无人机缓存博弈问题。尽管UAV拥有移动性的优势,但仍然受到一些实际条件的限制,例如禁飞区、安全问题等,无人机有限的电池容量也是需要重点考虑的因素之一。本文专注于无人机的缓存及部署问题研究,将在后续研究中进一步考虑有限电池容量的缓存问题。对于无人机作为基站的能源问题,可以采用系留无人机充当基站、利用太阳能充电、使用低功耗技术、应用无线功率传输技术等方式解决。此外,在应急场景中无人机是不需要长时间运行的。已有文献提出了一些无人机电源问题的解决方案,例如文献[24]提出的基于无线功率传输的方案,文献[25]提出的基于太阳能电池的解决方案,文献[26]提出的高能效的UAV通信架构解决方案。

在缓存优化问题中,对内容流行度的正确解读至关重要,因为其影响着缓存网络的部署及缓存策略的制定,这很大程度上决定了总体网络性能。然而,对内容流行度的理解、追踪及预测是复杂的,并且经常被错误解读。缓存研究中使用最广泛的流行度模型是独立参考模型(IRM, independent reference model),并假设内容请求服从独立同分布的概率分布,并且不随时间而改变<sup>[27]</sup>。该模型虽然使缓存问题模型易于处理,但牺牲了模型的准确性,尤其是在大时间尺度分析中是不合理的。内容流行度动态特征的重要性还体现在在线缓存策略的研究上。例如,最少使用(LRU, least recently used)策略和最不经常使用(LFU, least frequently used)策略通过调整缓存决策来适配内容流行度的波动。在线缓存策略虽然不能达到理论上的性能最优,但由于其易实现性及对非静态流行度的适配能力,在实际工程系统中非常实用<sup>[28]</sup>。不同于被动缓存策略,最近的研究工作大多关注预测内容的流行度,并基于流行度进行内容的最优化部署。例如,利用

社交网络预测内容流行度<sup>[29]</sup>、利用机器学习技术研究流行度的变化<sup>[30]</sup>等。

用户移动性对缓存部署算法的设计同样至关重要。文献[31]将用户移动特性分为2种，即空间特性和时间特性。空间特性包含与物理位置有关的用户行为模式的信息。时间特性描绘用户的时间属性。用户的行为模式通常具有一定的周期特性，在设计缓存部署算法时可将其考虑在内。用户的返回时间（即任意用户重新回到之前访问过的地点的时间）可以用来衡量移动用户重返某一区域的周期特性和频度。文献[32]基于一组大学校园的数据研究发现，移动用户的活动区域有限且拥有一定的周期性。文献[33]研究了用户返回时间的分布规律，结果显示，用户在24 h、48 h和72 h后返回原位的概率最大。

无线网络中的用户行为是动态的，因此用户内容请求在时间和空间上相应地显现一定的规律性。基于此，本文研究如何在这种动态的用户行为模式场景下，利用无人机的可移动特性及缓存能力，更好地为移动用户提供内容传输服务。

### 1.1 系统模型

本文研究的无人机缓存系统模型如图1所示。一个宏基站覆盖区域内共有 $U$ 个移动用户请求内容传输服务，所有用户表示为集合 $\mathcal{U}$ 。在本文模型中，部署 $K$ 个配置有缓存单元的无人机作为飞行的小基站，与宏基站协作来为地面用户提供内容传输服务，所有无人机表示为集合 $\mathcal{K}$ 。由于无人机的高度可调节性可以极大地降低因障碍物而引起的阻挡效应，因此本文研究无人机到地面用户的空地通信链路工作在毫米波频段。此外，无人机作为部署在空中的小基站，使用蜂窝网许可频段作为前传链路与宏基站相连。宏基站通过光纤回程链路与内容服务器连接。

在本文模型中，内容服务器共存储了 $N$ 个内容，用户向该 $N$ 个内容中的部分内容发起请求，所有内容表示为集合 $\mathcal{N}$ 。每个内容大小为 $H$  bit。无人机可以缓存用户将来可能请求的内容，从而直接将所缓存的内容传输给移动用户，从而显著减少传输时延。本文将无人机 $k(k \in \mathcal{K})$ 所缓存的内容表示为集合 $\mathcal{C}_k$ ，无人机的存储容量为 $CH$  bit，即最多可缓存 $C$ 个完整的内容，其中 $C < N$ 。假设缓存在无人机中的内容每隔一段时间更新一次，更新过程发生在网络的空闲时段，通常无人机在该时段返回停机地点完成充电等任务。

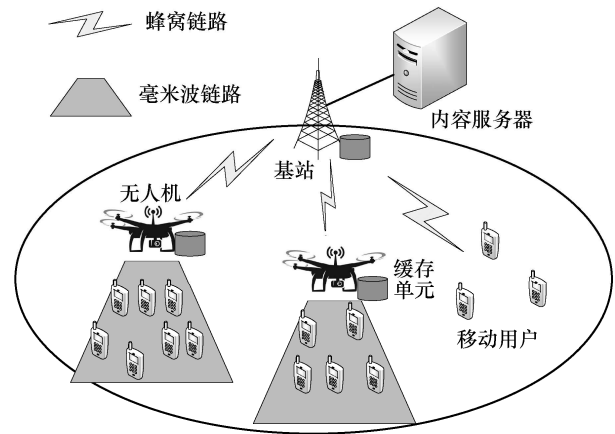


图1 无人机缓存系统模型

### 1.2 用户文件请求模型

用户文件请求模型对无人机缓存部署问题至关重要。用户文件请求模型主要包括用户对文件的偏好（即请求概率）、用户发起文件请求的时间和位置。

#### 1) 用户对文件的偏好

所有用户的文件请求概率矩阵 $\mathbf{P}$ 如式(1)所示。

$$\mathbf{P} = \begin{pmatrix} p_{11} & \cdots & p_{1N} \\ \vdots & \ddots & \vdots \\ p_{U1} & \cdots & p_{UN} \end{pmatrix} \quad (1)$$

其中， $p_{un}$ 为第 $u$ 行第 $n$ 列的元素，表示用户 $u$ 对文件 $n$ 的请求概率。

根据式(1)可知，文件 $n$ 的全局流行度 $P_n = \sum_{u=1}^U p_{un}$ ，表示大量用户请求文件 $n$ 的平均概率。此外， $P_n$ 可以在宏基站侧通过分析一个统计间隔内的用户行为来获得，例如，统计单位时间内每个文件被请求的次数、统计固定的总请求次数中每个文件被请求的次数。

#### 2) 用户发起文件请求的时间和位置

与文献[31]类似，本文假设用户发起文件请求的时间和位置是周期性的，并假设每个用户每天只发起一次文件请求，且在文件传输过程中位置保持不变。

用户发起文件请求的位置对UAV位置的部署起着决定性作用。根据用户发起文件请求的位置来部署UAV，能够最小化传输距离，从而节约能量、提升系统吞吐量。

### 1.3 问题构建

当网络繁忙时，用户的平均带宽低、传输速率低、时延大、回传链路负载高。此时通过部署无人

机, 可以将内容缓存在无人机中然后传输给用户, 这些内容可在网络闲时缓存到无人机中。这种边缘缓存策略在增强用户 QoE 的同时降低对回传链路容量的需求。

在本文模型中, 无人机作为可移动的小基站为用户提供服务, 可动态地调整位置, 从而满足用户的内容请求。

本文的优化目标为最大化无人机侧的缓存命中率, 限制条件为无人机有限的存储容量, 待优化变量为无人机的位置及无人机侧的内容部署。通过本文所提算法, 实现用户对文件的偏好、用户发起文件请求的时间和位置的预测。

本文将内容请求命中率定义为若某一时刻  $t$ , 用户  $u$  能够被其附近且缓存有相应内容的无人机  $k$  服务, 则其内容请求可被满足并归入总效用。总体缓存命中率为

$$h(\mathbf{F}, \mathbf{L}) = \sum_u \sum_n R_{un} \frac{F_{k(u)n}}{N_R} \quad (2)$$

其中,  $R_{un}$  表示在高峰时段用户  $u$  对文件  $n$  的请求指示, 用户  $u$  请求了文件  $n$ ,  $R_{un}=1$ , 否则  $R_{un}=0$ ;  $k(u)$  ( $k(u) \in \mathcal{K}$ ) 表示为用户  $u$  提供服务的无人机编号, 其取值与用户和 UAV 的位置有关;  $F_{kn}$  为无人机  $k$  缓存文件  $n$  的指示变量, 若无人机  $k$  缓存了内容  $n$ , 则  $F_{kn}=1$ , 否则  $F_{kn}=0$ ;  $N_R$  表示所有用户的内容请求总次数。

由于无人机的位置决定了其所能提供服务的用户集合, 影响了缓存命中率, 因此, 待优化变量为无人机内容部署决策矩阵  $\mathbf{F}$  和无人机的部署位置  $\mathbf{L}$ , 限制条件为在每个无人机缓存中存储的文件数量小于无人机配备的存储容量。因此, 优化问题可构建为式(3)所示形式。

$$\begin{aligned} \max_{\mathbf{F}, \mathbf{L}} & \sum_u \sum_n R_{un} \frac{F_{k(u)n}}{N_R} \\ \text{s.t.} & \sum_{n=1}^N F_{kn} \leq C \end{aligned} \quad (3)$$

式(3)为多优化目标的非线性规划问题, 传统解决方法的计算代价极高, 因此, 本文提出了启发式算法, 具体如下: 首先预测用户文件请求概率和请求时间及位置; 然后根据请求概率和用户发起请求的位置来决定无人机的部署位置, 并由无人机的位置决定缓存内容的部署。

## 2 内容请求概率和用户行为模式预测

本节提出了用户未来的内容请求概率分布和行为模式预测算法, 预测的结果将被用于无人机飞行轨迹及内容部署的设计。

### 2.1 单个用户文件偏好预测

变量  $p_{um}$  的预测对主动缓存优化问题至关重要。因为全局流行度体现的是大量用户集体内容的流行度趋势, 所以在缓存部署阶段使用单个用户的内容流行度比使用全局的内容流行度更加合理。

基于协同过滤<sup>[34]</sup>的算法在推荐系统中被广泛应用, 并且能够达到较好地预测用户需求的效果。因此, 本文所提的文件偏好预测算法同样基于协同过滤, 同时利用了用户的上下文信息加强对单个用户文件流行度预测的准确性。

#### 1) 基于文件相似性的内容请求预测

基于上下文信息的相关性来预测未来的内容流行度, 也就是根据用户的文件请求历史信息反映出的个人喜好来预测未来的内容请求概率。具体是根据目标文件和历史请求文件之间的相似性来预测流行度。

对于文件  $n$ , 本文用集合  $\{f_{n1}, f_{n2}, \dots, f_{nG}\}$  表示文件的类别, 每个元素表示一个分类。例如,  $f_{n1}$  表示记录片,  $f_{n2}$  表示喜剧片...当文件  $n$  属于该类别时, 值为 1, 否则为 0。下面定义文件  $n$  和文件  $m$  的相似度  $\text{Sim}_f(n, m)$ 。相似度的计算方法基于余弦相似性, 具体如式(4)所示。

$$\text{Sim}_f(n, m) = \frac{\sum_{g=1}^G (f_{ng} f_{mg})}{\sqrt{\sum_{g=1}^G (f_{ng})^2} \sqrt{\sum_{g=1}^G (f_{mg})^2}} \quad (4)$$

根据用户的历史文件请求信息和文件间的相似性, 可以计算出用户对于未请求过内容的偏好程度。例如, 用户  $u$  对文件  $n$  的请求概率值可通过式(5)来预测得到。

$$p_{um}^f = \sum_{m \in F_u} \text{Sim}_f(n, m) p_{um} \quad (5)$$

其中,  $\{m | m \in F_u\}$  表示用户  $u$  所请求过的历史文件子集合;  $p_{um}$  表示用户  $u$  对文件  $m$  的请求概率, 假设其与对文件  $m$  的平均请求概率  $P_m$  相等。在计算得到所有  $p_{um}^f$  后, 对其做归一化处理, 即可得到基

于文件相似性而预测的内容请求概率。

## 2) 基于用户相似性的内容请求预测

除了文件相似性,也可以基于用户的人口学信息相关性来预测内容流行度。具体地,拥有相似背景的用户,例如受教育水平、年龄等特征,即使没有社交关系也可能具有类似的偏好。也就是说,用户的偏好与性别、年龄、文化程度等信息是相关的。通过将用户分为不同的群体,得到用户之间的相关性,并用来加强对未来网络中的用户行为预测的准确性。

对于用户  $v$ , 用集合  $\{b_{v1}, b_{v2}, \dots, b_{vL}\}$  表示用户的不同特征。例如,  $b_{vl} (l \in L)$  可以分别表示年龄、性别、教育、职业、区域、收入等其他人口学特征。用户  $u$  和  $v$  的相似性通过计算皮尔逊相关系数获得, 如式(6)所示。

$$\text{Sim}_u(u, v) = \frac{\sum_{l=1}^L (b_{ul} - \bar{b}_u)(b_{vl} - \bar{b}_v)}{\sqrt{\sum_{l=1}^L (b_{ul} - \bar{b}_u)^2} \sqrt{\sum_{l=1}^L (b_{vl} - \bar{b}_v)^2}} \quad (6)$$

其中,  $b_{ul}$  和  $b_{vl}$  表示用户  $u$  和  $v$  的第一个属性值,  $\bar{b}_u$  和  $\bar{b}_v$  表示用户  $u$  和用户  $v$  所有属性的平均值。每个用户的属性向量取值用数字表示, 例如, 性别可表示为 0 或 1; 不同年龄的人群可以分类表示, 0~5 岁为 0, 6~10 岁为 1, 以此类推。

用  $U_v^M = \{u_1, u_2, \dots, u_M\}$  表示与用户  $v$  最相似的  $M$  个用户的集合, 本文将此集合归类为一个用户群体。因此, 考虑来自同一群体的用户对流行度的影响, 用户  $v$  对文件  $n$  的请求概率为  $p_{vn}^b$ , 可通过统计用户集合  $U_v^M$  中所有用户对预测目标内容的请求数量, 再计算每个预测目标内容的请求频率得到。

## 3) 综合考虑文件相似性和用户相似性

如上所述, 本节分别分析了文件相似性和用户相似性对内容请求概率的影响。为了综合考虑这两方面因素对预测的影响, 提出预测用户  $u$  对文件  $n$  的请求概率的模型, 如式(7)所示, 且假设 2 个因素  $p_{un}^f$  和  $p_{un}^b$  对于总体的内容请求概率预测值  $p_{un}$  的影响是线性的。

$$p_{un} = \alpha_1 p_{un}^f + \alpha_2 p_{un}^b \quad (7)$$

其中,  $\alpha_1$  和  $\alpha_2$  表示相应的权重系数, 满足等式  $\alpha_1 + \alpha_2 = 1$ , 可以通过对它们的不同组合进行实验决定取值。

## 2.2 用户发起文件请求的时间和位置预测

用户的行为模式可分为两类, 即时间上的行为模式和空间上的行为模式。由于本文的目标是最优化用户的文件请求命中率, 因此本文只关注与内容请求相关的移动用户行为模式预测。

考虑到大部分人日常活动的发生时间和发生地点具有规律性和稳定性, 因此本文利用线性回归的方法来预测用户发起文件请求的时间点和地理位置。

假设单个用户每天请求一次内容传输(对于单个用户每天多次请求内容传输的情况, 可以将多次的内容请求看成不同用户的请求), 可以预测用户  $u$  的下一请求发起时间点为

$$\hat{T}_u(t) = \sum_{\tau=1}^{t-1} \beta_\tau T_u(\tau) \quad (8)$$

其中,  $T_u(\tau)$  表示用户  $u$  的第  $\tau$  次发起请求的时间点,  $\beta_\tau$  表示与其相对应的参数。参数  $\beta_\tau$  的估计通过训练数据集来获得, 最常用的参数估计方法是最小二乘法。

在本文中, 内容请求时间点既是输入量也是输出量, 参数  $\beta_\tau$  可看作  $T_u(\tau)$  和  $\hat{T}_u(t)$  之间的归一化相关系数。假设过去的请求时间点会激励未来的时间点, 因此系数  $\beta_\tau$  是非负的。同时, 时间间隔较短的请求之间拥有较强的相关性, 即系数关于请求间隔是非增的。

基于以上分析,  $\beta_\tau$  可通过求解具有线性约束条件的最优化问题获得, 如式(9)所示。

$$\begin{aligned} \min_{\beta_\tau} & \sum_{d=1}^D (\hat{T}_u(d) - T_u(d))^2 \\ \text{s.t.} & \beta_\tau - \beta_{(\tau+1)} \geq 0, \quad \forall \tau \\ & \beta_\tau \geq 0, \quad \forall \tau \end{aligned} \quad (9)$$

式(9)所示问题是线性约束条件下的非线性规划问题, 可以利用梯度投影法求解。通过加入线性约束条件来求解  $\beta_\tau$ , 可以加快线性回归问题求解的收敛速度, 同时也防止过拟合问题的出现。

同理, 对于用户发起请求时的地理位置的横纵坐标值, 可以分别运用类似的线性回归方法预测得到。

## 3 无人机位置部署和内容缓存

### 3.1 无人机位置部署

本文根据用户的文件偏好和预测的用户发起

请求时的位置决定无人机部署的位置。首先考虑用户的文件偏好在无人机部署位置过程中的作用,即当相似文件偏好的用户由同一无人机提供服务时,通过无人机缓存可以达到较高的文件请求命中率。用户发起请求时的位置决定了无人机部署在到用户间的传输距离尽量短的位置。对于前者,本文采用基于自组织映射神经网络(SOM, self-organizing map)的聚类算法,将文件偏好相似的用户聚成一类;对于后者,本文利用基于凝聚嵌套(AGNES, agglomerative nesting)的分簇算法,将每个类内距离相近的用户分在一个簇内。最后,根据分簇结果决定为每个簇服务的无人机的部署位置。

1) 基于 SOM 的聚类算法

相比于受初始化和数据噪声影响比较大的  $K$  平均算法, SOM 算法不固定聚类数量,因此受初始化影响小,并且在更新参数时也会更新邻近节点参数。因此,本文使用基于 SOM 算法进行聚类。

芬兰赫尔辛基大学的 Kohonen<sup>[35-37]</sup>于 1981 年提出了自组织映射算法,有时也称之为 Kohonen 网络。该算法是无监督学习中一类重要的机器学习方法,可作为一种聚类方法,也可用于特征提取、数据压缩、高维可视化等。

本文所采用的 SOM 本质上是一个两层的神经网络,如图 2 所示。其中,输入层为预测的用户内容偏好,输出层中的神经元为对用户偏好的聚类结果,表示最终聚成的类。

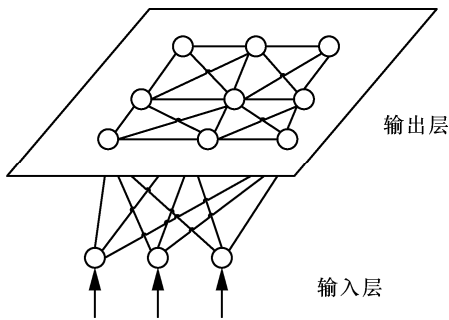


图 2 SOM 结构

图 2 所示神经网络以“竞争学习”的方式进行训练。每次输入一个用户的内容偏好,输出层中存在一个与该输入模式最匹配的聚类点,称为该输入模式的激活节点,也称为“获胜神经元”。每次训练后都需要调整激活节点及其邻近点的参数。

本文所提基于 SOM 聚类算法具体步骤如算法 1 所示。其中假设输入空间是  $N$  维,即将系统中

$U$  个用户对  $N$  个文件的内容偏好预测作为输入层上的  $N$  维输入模式,对应于用户  $u$  的输入模式可以表示为  $p^u = \{p_{u1}, \dots, p_{un}, \dots, p_{uN}\}$ 。输出层上的  $V$  个神经元将作为聚类后的  $V$  个用户类,表示为  $Y = \{Y_1, \dots, Y_v, \dots, Y_V\}$ 。输入单元  $n$  和神经元  $v$  之间在计算层的连接权重表示为  $w_{vn}$ 。另外,  $S_{m,Y(p^u)}$  表示输出层的获胜神经元  $Y(p^u)$  和任意神经元  $m$  之间的距离。

$\sigma(t)$  的值随时间而衰减,  $\sigma(t) = \sigma_0 \exp\left(-\frac{t}{\tau_\sigma}\right)$ 。

学习率  $\eta(t)$  是以时间  $t$  为变量的函数,

$$\eta(t) = \eta_0 \exp\left(-\frac{t}{\tau_\eta}\right)$$

算法 1 基于 SOM 的聚类算法

**初始化** 聚类数量  $V=U$ , 输入单元  $n$  和神经元  $v$  之间的连接权重  $w_{vn}$  为  $0 \sim 0.01$  的随机数。

**输出** 每个  $p^u = \{p_{u1}, \dots, p_{un}, \dots, p_{uN}\}$  的获胜神经元为最终聚类结果

① while 特征映射不稳定即获胜神经元仍在变化

② for  $u=1:U$

③ 用户  $u$  的输入模式  $p^u = \{p_{u1}, \dots, p_{un}, \dots, p_{uN}\}$

④ for  $v=1:V$

⑤ 计算每个神经元  $Y_v$  的判别函数值

$$d_v(p^u) = \sum_{n=1}^N (p_{un} - w_{vn})^2$$

⑥ end for

⑦ 确定  $d_v(p^u)$  最小的神经元  $Y(p^u)$  为获胜神经元

⑧ for  $v=1:V$

⑨ 计算获胜神经元  $Y(p^u)$  与兴奋神经

$$\text{元拓扑领域的空间位置 } T_{v,Y(p^u)}(t) = \exp\left(-\frac{S_{v,Y(p^u)}^2}{2\sigma(t)^2}\right),$$

并相应地调整相关兴奋神经元的连接权重数值

$$\Delta w_{vn} = \eta(t) T_{v,Y(p^u)}(t) (p_{un} - w_{vn})$$

⑩ end for

⑪ end for

⑫ end while

2) 基于 AGNES 的分簇算法

本文采用基于 AGNES 的分簇算法,根据用户的位置对每个用户内容偏好类进行分簇,便于在最

佳位置部署无人机。考虑到无人机部署的特性，层次聚类算法<sup>[38-41]</sup>更加适合本文场景，该类算法试图在不同层次对数据集进行划分，最终形成一个树型的聚类结构。

基于 AGNES 的分簇算法是一种采用自底向上聚合策略的层次聚类算法<sup>[42]</sup>。基于 AGNES 的分簇算法将数据集中的每个样本看作一个初始聚类簇，然后在算法运行的每一步中找出距离最近的 2 个聚类簇进行合并，该过程不断重复，直至达到预设的聚类簇距离阈值。本文考虑以聚簇类距离阈值作为算法迭代停止条件，原因是本节考虑无人机到用户间传输覆盖距离，这样能够获得不固定的分簇数量，实现无人机部署数量的自适应调整，从而提高无人机部署效率，减小无人机间干扰。

以基于 SOM 聚类算法后的第  $q$  类中  $Q$  个用户进行分簇为例，基于 AGNES 的分簇算法步骤如算法 2 所示。

**算法 2** 基于 AGNES 的分簇算法

输入 样本集  $D_q = \{x_1, x_2, \dots, x_Q\}$

输出 簇划分  $\{C_1, C_2, \dots, C_k\}$

① for  $j = 1, 2, \dots, Q$  do

②  $C_j = \{x_j\}$

③ end for

④ for  $i = 1, 2, \dots, Q$  do

⑤ for  $j = i+1, \dots, Q$  do

⑥  $M(i, j) = d(C_i, C_j)$

⑦  $M(j, i) = M(i, j)$

⑧ end for

⑨ end for

⑩ 设置迭代停止聚类簇距离阈值  $\eta$

⑪ while  $\min(M(i, j)) > \eta$  do

⑫ 找出距离最近的 2 个聚类簇  $C_i$  和  $C_j$  ( $i^* < j^*$ )

⑬ 合并  $C_i$  和  $C_j$ :  $C_{i^*} = C_i \cup C_j$

⑭ for  $j = j^*+1, \dots, q$  do

⑮ 将聚类簇  $C_j$  重新编号为  $C_{j-1}$

⑯ end for

⑰ 删除距离矩阵  $M$  的第  $j^*$  行与第  $j^*$  列

⑱ for  $j = 1, 2, \dots, q-1$  do

⑲  $M(i^*, j) = d(C_{i^*}, C_j)$

⑳  $M(j, i^*) = M(i^*, j)$

㉑ end for

㉒  $Q = Q - 1$

㉓ end while

分簇算法的关键是如何计算聚类簇之间的距离。事实上，每个簇是一个样本集合，因此只需要采用关于集合的某种距离即可。例如给定聚类簇  $C_i$  和  $C_j$ ，可通过式(10)来计算距离。

$$d(C_i, C_j) = \max_{x \in C_i, z \in C_j} \|x - z\| \quad (10)$$

其中， $\|x - z\|$  表示聚类簇  $C_i$  和聚类簇  $C_j$  内的点  $x$  和  $z$  之间的平面距离。

### 3) 选择无人机位置

本文基于 AGNES 分簇算法对每类用户进行分簇，选择最佳的无人机部署位置来向每一分簇内用户分发内容，以达到减少系统总能耗、缓解基站负载压力的目标。

假设在一个含有  $S$  个用户的簇  $C_s$  中， $\{y_1, \dots, y_s, \dots, y_S\}$  表示簇内用户位置坐标。本文将服务于该用户簇的无人机部署位置选择问题构建为式(11)所示的优化问题，即最小化无人机到该簇内所有用户距离之和。

$$\min_x \sum_{s=1}^S \|x - y_s\| \quad (11)$$

其中，无人机部署位置平面坐标为  $x = (a, b)$ ，用户  $s$  坐标表示为  $y_s = (a_s, b_s)$ ， $\|x - y_s\|$  为无人机和用户  $s$  的平面距离。可以较容易证明该问题为凸优化问题，并且可采用梯度下降法求解最优无人机部署平面坐标值  $x^* = (a^*, b^*)$ 。

### 3.2 无人机缓存内容部署

移动网络运营商的用户基站连接策略在缓存部署设计中至关重要。一方面，信号强度影响着无线传输的性能。另一方面，当用户连接到未缓存所请求内容的基站时，缓存增益为零。

传统的蜂窝网用户连接策略只考虑信号质量，即将蜂窝用户与地理上最近的基站连接（简称“最近无人机用户链接策略”）。在考虑最大化缓存增益的场景中，用户会优先接入能够满足一定信号质量并且缓存了所请求内容的基站当中（简称“缓存感知用户连接策略”）。本节分别讨论了以上 2 种不同用户基站连接策略下的缓存部署算法，具体介绍如下。

### 1) 最近无人机用户连接策略

在此用户基站连接策略下, 即当用户在无人机的覆盖范围内时, 由最近的无人机提供服务, 否则由宏基站提供服务。基于无人机的部署位置, 可以确定用户与无人机在某个时刻的连接拓扑。

具体地, 移动用户对于服务无人机的选择原则如式(12)所示。

$$k(u) = \arg \min_k (\|l_u - l_k^{\text{uav}}\|) \quad (12)$$

式(12)表示用户终端选择距离最近的无人机为其提供服务。其中,  $l_u$  表示用户  $u$  的笛卡尔位置坐标,  $l_k^{\text{uav}}$  表示无人机  $k$  的位置坐标。

根据式(12)所示的用户与无人机的连接规则, 用户在无人机覆盖范围内时, 与最近的无人机通信, 因此无人机之间的内容部署不存在耦合关系。相应地, 只需分别优化每个无人机所服务用户的命中率。优化问题可退化为如式(13)所示的问题。

$$\begin{aligned} \mathbf{F}(k,:) &= \arg \max_{\mathbf{F}(k,:)} \sum_{n=1}^N F_{kn} p_n^k \\ \text{s.t.} \quad & \sum_{n=1}^N F_{kn} \leq C \end{aligned} \quad (13)$$

其中,  $\mathbf{F}(k,:)$  为无人机  $k$  的内容部署向量,  $p_n^k$  为无人机  $k$  所服务的用户簇对于文件  $n$  的平均请求概率。

$$p_n^k = \frac{\sum_{u \in \mathcal{U}_k} p_{un}}{|\mathcal{U}_k|} \quad (14)$$

其中,  $\mathcal{U}_k$  表示无人机  $k$  所服务的所有用户的集合,  $|\mathcal{U}_k|$  表示该集合的势。

针对式(13)所示的问题, 本文设计无人机  $k$  的内容部署策略如下: 对无人机  $k$  服务的所有用户的内容偏好进行统计平均, 得到一个总体内容请求概率向量, 取该向量中流行度最高的  $C$  个内容缓存到无人机  $k$  中即可。

### 2) 缓存感知用户连接策略

当用户优先接入缓存了所请求内容的无人机, 并且在已知无人机的位置情况下, 与文献[43]类似, 式(1)所示优化问题可被重新构建为一个在拟阵约束条件下的单调次模函数最优化问题。该问题是 NP 完全问题, 解决该问题的常用方法是贪婪算法。贪婪算法以一个空集作为初始的优化变量, 在每次迭代中, 在该集合中加入能够为目标函数带来最大

边际增量的元素, 同时保持解的独立性。文献[44]的结果表明, 贪婪算法可以达到最优解的  $\frac{1}{2}$ 。基于贪婪算法的缓存部署伪代码如算法 3 所示。

#### 算法 3 基于贪婪算法的缓存部署

**初始化** 将缓存部署集合  $X$  设为空集, 即  $X = \emptyset$

**输出** 缓存部署集合  $X$

① while 无人机的缓存空间未满足时

② for  $j=1:NK$

③  $h_x(j) = h(X + j) - h(X)$

④ end for

⑤  $F_k^n = \arg \max_j h_x(j)$

⑥ 将  $F_k^n$  加入集合  $X$  中, 即将内容  $n$  部署到无人机  $k$  的缓存当中

⑦ end while

至此, 本文已完成了无人机的位置和内容部署与动态的用户内容请求的适配, 其中, 算法 1 的复杂度为  $O(2NU^2 + U^3)$ , 算法 2 的复杂度为  $O(U + 3U^2)$ , 算法 3 的复杂度为  $O(3NK)$ 。本文方案的复杂度分析如下。

1) 虽然文件内容是不断更新的并且文件总数较大, 可能会造成方案的复杂度过大, 但是由于长尾效应, 只有较少的文件是最流行的, 本文考虑的文件偏好也是针对这部分最流行内容, 因此内容数量  $N$  不会对方案复杂度造成太大影响。同时, 这些最流行内容可由内容供应商的服务器统计获得, 由于在内容供应商服务器处的信息集中且计算能力强, 因此能够以较快的速率、较高的精度找到最流行内容, 并且复杂度是可以忽略的。

2) 用户发起请求时的位置虽然也在动态变化, 但一般波动较小, 可认为是千米级的, 因此用户呈现出一定的社区性, 而无线网络中千米级覆盖范围内的用户数量  $U$  及无人机数量  $K$  均不会特别大。此外, 随着软件定义网络、网络功能虚拟化、边缘计算等技术的发展, 并且考虑到本文所提方案是在网络空闲时段执行的, 周期按天计算, 故方案复杂度可认为是网络能够负担的。换句话说, 本文方案的复杂度是可接受的。

## 4 仿真结果

首先, 本文仿真并分析了预测算法的准确性。在评估预测算法准确性时, 由于真实的用户历史信息由无线网络运营商掌握, 很难获得, 因此本文使

用的数据集为公共的 MovieLens 数据集<sup>[45]</sup>。其次，本文分别对缓存命中率和传输时延受不同系统参数变化的影响进行了实验仿真与分析，仿真结果采用蒙特卡洛法多次重复执行算法，然后取平均值作为结果。

#### 4.1 内容请求预测准确度

本文所使用的 MovieLens 数据集由 GroupLens 组织提供，并且已经被广泛应用于教育、研究和工业领域的科研和实践当中。本文需要通过用户的人口学信息来预测内容请求概率，因此选择使用 MovieLens 数据集的 100k 版本。该版本完整的数据集包含了 943 名用户对 1 682 部电影的共计 100 000 条评分信息，每个用户至少对 20 部电影进行了评分。本文认为一个电影评分意味着在该时间点向无线网络发出的一个对该内容的请求。根据长尾理论，只有很少一部分的内容被大多数的用户请求。因此本文仿真中仅对请求概率最高的前 100 部电影进行了实验分析。此外，该数据集也包含了用户的人口学信息，可以用来计算用户相似性。部分 MovieLens 数据集源数据（u.data 文件和 u.user 文件）如表 1 和表 2 所示。表 1 数据包括用户 ID、电影 ID、评分和时间戳，其中，用户 ID 和电影 ID 可用来预测用户对内容的偏好，评分表示用户对电影的打分，时间戳表示评分的时间。本文忽略评分高低，仅认为其代表该时间点向无线网络发出的内容请求。表 2 数据包括用户 ID、年龄、性别、职业等信息，可用来分析用户之间的相似性。

表 1 u.data 文件的源数据（部分）

用户 ID	电影 ID	评分	时间戳
196	242	3	881 250 949
186	302	3	891 717 742
22	377	1	878 887 116
244	51	2	880 606 923

表 2 u.user 文件的源数据举例（部分）

用户 ID	年龄/岁	性别	职业
1	24	男	技师
10	53	男	律师
15	49	女	教师
23	30	女	艺术家

历史请求数据量对用户内容偏好预测准确性的影响如图 3(a)所示。历史请求数据量是在某一时间点用户已评分过的所有电影数量，准确度为预测

的偏好向量中请求概率最大的文件在未来被请求的概率，即未来被请求次数除以总的预测次数。可以看出，随着用户的历史请求文件数量上升，用户偏好的预测更加准确。也就是说，通过文件相似性的计算，历史请求文件越多，就可以越准确地估计用户对某一类型电影的偏好程度。

不同特征的用户之间内容偏好的相似性如图 3(b)所示，由此可分析出内容偏好预测模型中用户特征的影响。仿真中分别计算了不同年龄差用户之间的流行度余弦相似性值，以及某个职业用户与其他非该职业用户流行度余弦相似性值。可以看出，年龄差距越大，不同年龄段用户的内容流行度相似度越低，也就是内容偏好差异越大。另一方面，相比较年龄差距，不同职业用户群体之间的内容偏好差异较大。也就是说职业类型与用户偏好间的相关性较年龄差距的相关性大。因此在利用用户相似性预测单个用户的内容偏好时，相比于年龄差距特征对用户偏好的影响程度，应该为职业类型特征赋予更大的影响权值。

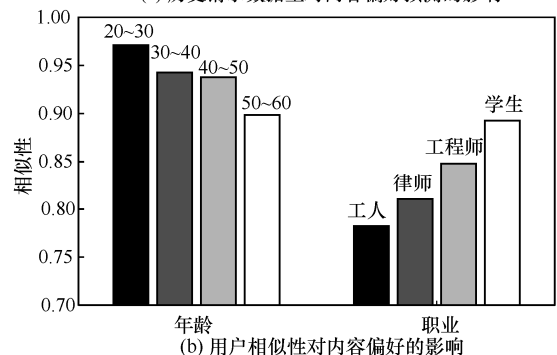
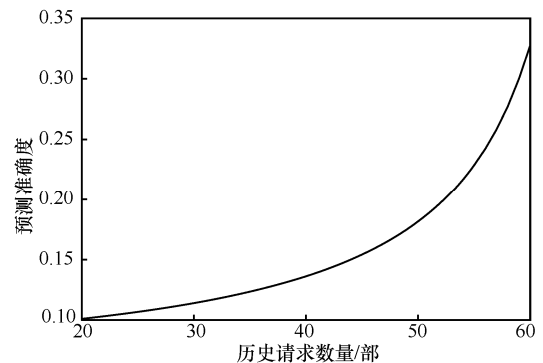


图 3 用户内容偏好预测仿真结果

由于 MovieLens 数据集中的时间信息和位置信息不适合验证本文提出的针对请求时间和位置的预测算法的合理性，故本文采用 Brightkite 数据集<sup>[46]</sup>，其数据举例如表 3 所示，其中每行数据分别为用户

ID、check-in 时间及当时所处位置的经纬度。由于篇幅限制，本文仅列举了 ID 为 0 的用户的 4 天数据，其余的大量数据所呈现的规律与表 3 数据相似。从表 3 可以看出，用户的 check-in 时间较为固定，前后相差不超过 2 个小时。相对于缓存部署的时间尺度，这对算法设计造成的影响很小。此外，在同一时间用户的 check-in 经纬度也几乎相同，这些现象都证实了基于线性回归的预测算法的合理性。

**表 3 Brightkite 数据集数据举例**

用户 ID	check-in 时间	纬度	经度
0	2010-10-14 18:25:51	39.750 469°	-104.999 073°
0	2010-10-13 20:05:43	39.751 300°	-105.000 121°
0	2010-10-12 19:56:49	39.749 934°	-105.000 017°
0	2010-10-07 18:08:17	39.749 955°	-104.999 931°

此外，为了验证对于线性回归预测模型中线性系数的合理性，即时间上更近的数据值与预测值的相关性更大，本文根据 Brightkite 数据集中前 100 个用户的 check-in 位置，得到同一用户的请求位置距离与相隔天数的关系，如图 4 所示。

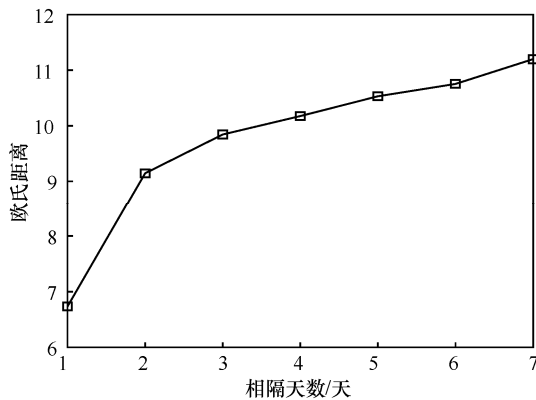


图 4 请求位置距离与相隔天数关系

图 4 中，相隔天数表示 2 个位置数据相隔的天数，请求位置距离为 2 个位置经纬度坐标值之间的欧氏距离。可以看出，相隔天数越短，2 个位置之间的欧氏距离越小，这也证明了在本文线性回归模型中对于时间上更近的数据值与预测值的相关系数更大的限制条件的合理性。本节仿真只给出最多相隔 7 天的数值，基于现实生活场景，本文认为大于 7 天的数值与预测值无相关性。此外，从图 4 中可看出，请求位置距离数值偏大，这与表 3 数据不符。其原因是数据中会包括相邻 2 天位置距离较大的情况，可以对应到现实生活中用户出差或者旅游等问

题中，但大部分时间的用户位置较稳定。因此，本文的线性回归模型无法准确预测位置突变的情况，但仍然具有较高的平均预测准确度。

#### 4.2 缓存命中率

缓存命中率和平均传输时延仿真所使用的通用参数设置如表 4 所示。

**表 4 通用参数设置**

符号	意义	数值
$U$	用户数量/个	30
$N$	文件数量/个	100
$K$	无人机数量/架	6
$C$	存储容量/个	10
$A$	地理区域	500 m×500 m
$H_{macro}$	宏基站高度/m	50
$H_{uav}$	无人机飞行高度/m	20
$D_{thres}$	簇簇距离阈值/m	200

相比于大部分相关文献忽略用户文件偏好和请求时间与地点对缓存算法性能的影响，本文的创新点体现在设计缓存算法时综合考虑这 3 个方面的影响。SOM 聚类算法针对用户文件偏好的设计考虑，不规定聚类数量，这对于不确定偏好类型数量的本文问题较有益。而基于 AGNES 的分簇算法则针对用户发起请求时的位置的设计考虑，该算法以距离作为阈值，这有助于无人机的高效覆盖。为了体现在设计缓存算法时将用户文件偏好和请求时间与地点考虑在内的益处，本节仿真了以下几种算法。

1) 本文算法 (分簇算法+本地流行度)。无人机的部署位置通过分簇算法获得；内容部署策略基于内容请求概率服从本地流行度的假设，即每个无人机服务的所有用户的平均内容流行度。

2) 对比算法 A (分簇算法+全局流行度)。无人机的部署位置通过分簇算法获得；内容部署策略基于内容请求概率服从全局流行度的假设，即整个仿真区域内所有用户的平均内容流行度。通过对比算法 A 与本文算法的对比分析，能够显示考虑内容偏好的增益，即基于 SOM 的聚类算法的潜在增益。

3) 对比算法 B (分簇算法+公平内容部署)。无人机的部署位置通过分簇算法获得；缓存策略为均匀地将内容部署在无人机侧，相当于随机缓存算法。

4) 对比算法 C (均匀无人机位置+全局流行度)。无人机的位置为固定且均匀分布的；内容部

署策略基于内容请求概率服从全局流行度的假设，即整个仿真区域内所有用户的平均内容流行度。通过对比算法 C 与对比算法 A 的对比分析，能够显示考虑请求位置的增益，即基于 AGNES 的分簇算法的潜在增益。

缓存命中率随着无人机的存储容量变化的情况如图 5(a)所示。可以看出，本文算法的缓存命中率性能优于其他对比算法。缓存命中发生有 2 个条件，即用户需要在无人机覆盖范围之内，并且无人机中缓存了相应的请求内容。本文算法通过分簇算法决定无人机位置，因此用户接入无人机概率较高。另一方面，本文算法考虑了用户的内容偏好，因此流行度估计更加准确。综合考虑，本文算法缓存命中率性能最佳。

图 5(b)为缓存命中率随着无人机数量变化的情况。可以看出，本文算法的缓存命中率随着无人机数量的增加而快速上升，其原因是随着无人机数量增加，每个无人机服务的用户数在减少，因此考虑了用户内容偏好的本地流行度估计用于缓存部署更加准确。同时，无人机数量的增多也会提高用户接入无人机的概率。基于以上 2 个因素，缓存命中率总体是增加的。而对比较算法的场景中，缓存命中率增加不明显的原因在于，无人机的数量增加虽然可以提高用户接入无人机的概率，却并不能提高缓存部署准确性，也就是没有基于 SOM 的聚类算法所代表的潜在增益，因此并不能带来缓存命中率的明显提升。

### 4.3 平均传输时延

本节假设用户的文件请求可以通过以下 2 种方式传输：方式(a)为内容服务器-宏基站-用户，方式(b)为无人机-用户。其中，当用户覆盖范围内没有无人机可接入，或者有无人机覆盖但并没有缓存所请求内容时，通过方式(a)来传输文件；当用户覆盖范围内有无人机可接入时，并且所请求文件被缓存在所接入的无人机存储器中，通过方式(b)来传输文件。

假设光纤连接宏基站和内容服务器作为回程链路，其产生的时延可忽略。因此，在时段  $t$  用户  $u$  接收文件  $n$  的时延为

$$D_{t,u,n}^a = \frac{L}{C_{BS-UE}} \quad (15)$$

$$D_{t,u,n}^b = \frac{L}{C_{CACHE-UE}} \quad (16)$$

其中， $C_{BS-UE}$  表示方式(a)下宏基站到用户链路的信道容量， $C_{CACHE-UE}$  表示方式(b)下缓存有用户所请求内容的无人机到该用户链路的信道容量。

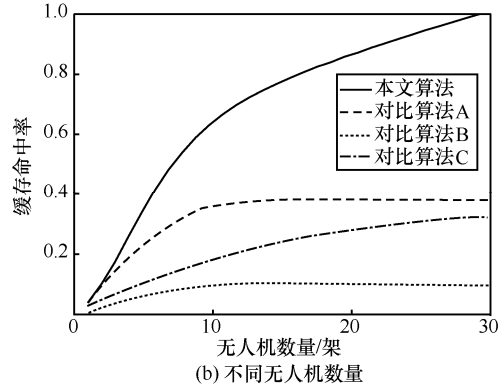
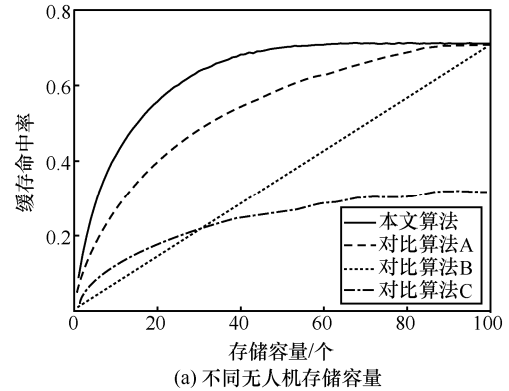


图5 缓存命中率

在仿真中，蜂窝频段带宽为  $W_1 = 10$  MHz，毫米波频段带宽为  $W_2 = 100$  MHz。宏基站发射功率为  $P_1 = 30$  dBm，无人机发射功率为  $P_2 = 20$  dBm。高斯白噪声功率谱密度为  $-174$  dBm/Hz。单个文件大小  $L$  为 100 MB。

图 6(a)为平均用户传输时延随着无人机缓存容量变化的情况。可以看出，本文算法时延性能优于其他对比算法。因为在考虑内容偏好时，基于 SOM 的聚类算法能够使流行度估计更加准确，缓存命中率更高，所以平均传输时延最低。另一方面，对比算法 A 的平均传输时延要比对比算法 C 的低，原因在于当使用分簇算法设计无人机的部署位置时，用户与无人机的平均距离更短，信道容量更高，从而时延更小。

图 6(b)为平均用户传输时延随着无人机数量变化的情况。可以看出，本文算法的时延随着无人机数量的增加而快速下降。随着无人机数量增加，每个无人机服务的用户数在减少，考虑了用户内容偏好的流行度估计用于缓存部署更加准确，缓存命中率提升，体现了基于 SOM 的聚类算法能够带来的

潜在增益。同时,用户平均接入带宽增加,信道容量提高,因此平均用户传输时延降低。而对比算法的场景中,时延下降不明显的原因在于无人机数量的增加并不能提高缓存部署的准确性,因此并不能够带来缓存命中率明显上升,故平均用户传输时延没有明显减小。尽管增加无人机数量可提高系统缓存命中率及时延性能,但也意味着部署成本的增加,因此在决定无人机部署数量时需要在两者之间折中选择。

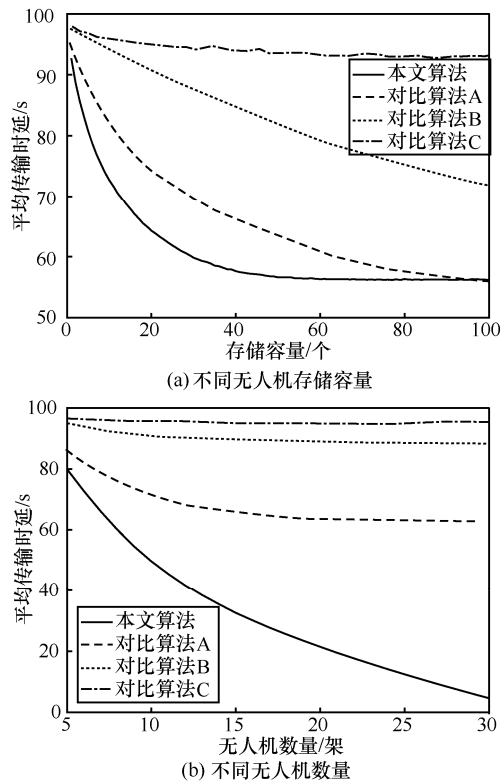


图 6 平均用户传输时延

## 5 结束语

为了适配用户的动态内容请求模式,提升缓存性能,本文讨论了无人机侧缓存部署策略,提出了基于文件相似性和用户相似性的单个用户内容偏好预测算法、用户发起请求时的位置和时间预测算法,并根据预测的结果利用聚类算法决定无人机的部署位置和内容部署策略。仿真结果表明,本文所提算法对内容流行度估计更加合理,同时,使用聚类算法拉近了用户与无人机之间的距离,因此在缓存命中率和平均传输时延方面的性能均超过了对比算法。此外,本文通过对数据集的分析观察到,不同职业用户群体之间的用户内容偏好差异较大。

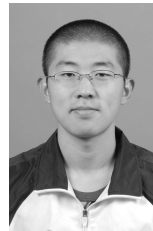
因此,在利用用户相似性加强内容偏好预测准确性时,应该为职业特征赋予更大的影响权值。

## 参考文献:

- [1] WANG X F, CHEN M, TALEB T, et al. Cache in the air: exploiting content caching and delivery techniques for 5G systems[J]. *IEEE Communications Magazine*, 2014, 52(2): 131-139.
- [2] 张平, 陶运铮, 张治. 5G 若干关键技术评述[J]. *通信学报*, 2016, 37(7): 15-29.  
ZHANG P, TAO Y Z, ZHANG Z. Survey of several key technologies for 5G[J]. *Journal on Communications*, 2016, 37(7): 15-29.
- [3] 张平, 牛凯, 田辉, 等. 6G 移动通信技术展望[J]. *通信学报*, 2019, 40(1): 141-148.  
ZHANG P, NIU K, TIAN H, et al. Technology prospect of 6G mobile communications[J]. *Journal on Communications*, 2019, 40(1): 141-148.
- [4] MOHAMMAD M A, DENIZ G. Fundamental limits of coded caching: improved delivery rate-cache capacity tradeoff[J]. *IEEE Transactions on Communications*, 2017, 65(2): 806-815.
- [5] LI B, FEI Z S, ZHANG Y. UAV communications for 5G and beyond: recent advances and future trends[J]. *IEEE Internet of Things Journal*, 2019, 6(2): 2241-2263.
- [6] ARMIR B, PALAZZI C E, RONZANI D. A comparison of stateless position-based packet routing algorithms for FANETs[J]. *IEEE Transactions on Mobile Computing*, 2018, 17(11): 2468-2482.
- [7] ZHANG S W, ZENG Y, ZHANG R. Cellular-enabled UAV communication: a connectivity-constrained trajectory optimization perspective[J]. *IEEE Transactions on Communications*, 2019, 67(3): 2580-2604.
- [8] CHENG N, XU W C, SHI W S, et al. Air-ground integrated mobile edge networks: architecture, challenges, and opportunities[J]. *IEEE Communications Magazine*, 2018, 56(8): 26-32.
- [9] 3GPP. New SID on enhanced support for aerial vehicles: RP-170779[R]. (2017-03-06)[2019-10-17].
- [10] 3GPP. Study on enhanced LTE support for aerial vehicles: TR36.777-f00[R]. (2017-12)[2019-10-17].
- [11] 3GPP. New WID on enhanced LTE support for aerial vehicles: RP-172826[R]. (2017-12-18)[2019-10-17].
- [12] 3GPP. New WID on UE conformance test aspects – enhanced LTE support for aerial vehicles: RP-182324[R]. (2018-12-10)[2019-10-17].
- [13] 3GPP. Draft New WID on NR enhancements for UAV: RP-192451[R]. (2019-12-09)[2020-04-14].
- [14] 3GPP. email discussion summary on Rel-17 NR UAV: RP-192452[R]. (2019-12-09)[2020-04-14].
- [15] 3GPP. discussion on NR-UAV for Rel-17: RP-192578[R]. (2019-12-09)[2020-04-14].
- [16] ZHANG H, SONG L, HAN Z, et al. Cooperation techniques for a cellular internet of unmanned aerial vehicles[J]. *IEEE Wireless Communications*, 2019, 26(5): 1-7.
- [17] FOTOUHI A, QIANG H, DING M, et al. Survey on UAV cellular communications: practical aspects, standardization advancements, regulation, and security challenges[J]. *IEEE Communications Surveys & Tutorials*, 2019, 21(4): 3417-3442.
- [18] MOZAFFARI M, SAAD W, BENNIS M, et al. A tutorial on UAVs for wireless networks: applications, challenges, and open problems[J].

- IEEE Communications Surveys & Tutorials, 2019, 21(3): 2334-2360.
- [19] BOR-YALINIZ I, SALEM M, SENERATH G, et al. Is 5G ready for drones: a look into contemporary and prospective wireless networks from a standardization perspective[J]. IEEE Wireless Communications, 2019, 26(1): 18-27.
- [20] CHEN M, MOZAFFARI M, SAAD W, et al. Caching in the sky: proactive deployment of cache-enabled unmanned aerial vehicles for optimized quality-of-experience[J]. IEEE Journal on Selected Areas in Communications, 2017, 35(5): 1046-1061.
- [21] XU X, ZENG Y, GUAN Y L, et al. Overcoming endurance issue: UAV-enabled communications with proactive caching[J]. IEEE Journal on Selected Areas in Communications, 2018, 36(7): 1231-1244.
- [22] ZHAO N, CHENG F, YU F R, et al. Caching UAV assisted secure transmission in hyper-dense networks based on interference alignment[J]. IEEE Transactions on Communications, 2018, 66(5): 2281-2294.
- [23] FANG T, TIAN H, ZHANG X, et al. Context-aware caching distribution and UAV deployment: a game-theoretic approach[J]. Applied Sciences, 2018, 8(10): 1-15.
- [24] LI L, XU Y, ZHANG Z, et al. A prediction-based charging policy and interference mitigation approach in the wireless powered Internet of things[J]. IEEE Journal on Selected Areas in Communications, 2019, 37(2): 439-451.
- [25] SUN Y, XU D, NG D W K, et al. Optimal 3D-trajectory design and resource allocation for solar-powered UAV communication systems[J]. IEEE Transactions on Communications, 2019, 67(6): 4281-4298.
- [26] ZENG Y, ZHANG R. Energy-efficient UAV communication with trajectory optimization[J]. IEEE Transactions on Wireless Communications, 2017, 16(6): 3747-3760.
- [27] BRESLAU L, CAO P, FAN L, et al. Web caching and Zipf-like distributions: evidence and implications[C]//International Conference on Computer Communications. Piscataway:IEEE Press,1999:1.
- [28] HASSLINGER G, NTOUGIAS K, HASSLINGER F, et al. Performance evaluation for new Web caching strategies combining LRU with score based object selection[J]. Computer Networks, 2017(125): 172-186.
- [29] HE S, TIAN H, LYU X C. Edge popularity prediction based on social-driven propagation dynamics[J]. IEEE Communications Letters, 2017, 21(5): 1027-1030.
- [30] HOU L, LEI L, ZHENG K, et al. A Q-learning based proactive caching strategy for non-safety related services in vehicular networks[J]. IEEE Internet of Things Journal, 2019, 6(3): 4512-4520.
- [31] WANG R, PENG X, ZHANG J, et al. Mobility-aware caching for content-centric wireless networks: modeling and methodology[J]. IEEE Communications Magazine, 2016, 54(8): 77-83.
- [32] RHEE I. CRAWDAD dataset[DB]. (2009-06)[2019-10-17].
- [33] GONZALEZ M C, HIDALGO C A, BARABASI A L. Understanding individual human mobility patterns[J]. Nature, 2008, 453(7196): 779-782.
- [34] LEE J, SUN M X, LEBANON G. A comprehensive study of collaborative filtering algorithms[J]. Cornell University, arXiv: Dos.3193. (2012-05-14)[2019-10-17].
- [35] KOHONEN T. Self-organized formation of topologically correct feature maps[J]. Biological Cybernetics, 1982, 43(1): 59-69.
- [36] KOHONEN T. An introduction to neural computing[J]. Neural Networks, 1988, 1(1): 3-16.
- [37] KOHONEN T. Self-organizing maps[M]. 3rd ed. Berlin: Springer, 2001.
- [38] KAUFMAN L, ROUSSEUW P J. Clustering by means of medoid[M]. Amsterdam: Elsevier, 1987.
- [39] KAUFMAN L, ROUSSEUW P J. Finding groups in data: an introduction to cluster analysis[M]. New York: John Wiley & Sons, 1990.
- [40] ZHANG T, RAMAKRISHNAN R, LIVNY M. BIRCH: an efficient data clustering method for very large databases[C]//Proceedings of the ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 1996: 103-114.
- [41] GUHA S, RASTOGI R, SHIM K. ROCK: a robust clustering algorithm for categorical attributes[C]//Proceedings of the 15th International Conference on Data Engineering. Piscataway: IEEE Press, 1999: 512-521.
- [42] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- ZHOU Z H. Machine learning[M]. Beijing: Tsinghua University Press, 2016.
- [43] GOLREZAEI N, SHANMUGAM K, DIMAKIS A G, et al. Femto-Caching: wireless video content delivery through distributed caching helpers[C]// International Conference on Computer Communications. Piscataway: IEEE Press, 2012: 1107-1115.
- [44] NEMHAUSER G L, WOLSEY L A, FISHER M L. An analysis of approximations for maximizing submodular set functions I[J]. Mathematical Programming, 1978, 14(1): 265-294.
- [45] MAXWELL H F, JOSEPH A K. The MovieLens datasets: history and context[J]. ACM Transactions on Interactive Intelligent Systems, 2015, 5(4): 1-19.
- [46] CHO E, MYERS S A, LESKOVEC J. Friendship and mobility: user movement in location-based social networks[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2011: 1-9.

## [作者简介]



任佳智 (1987- )，男，吉林省吉林市人，北京邮电大学博士生，主要研究方向为边缘缓存、网络切片、复杂网络。

田辉 (1963- )，女，河南郑州人，博士，北京邮电大学教授、博士生导师，主要研究方向为自组织网络、无线资源管理。

范绍帅 (1987- )，男，山东烟台人，博士，北京邮电大学讲师，主要研究方向为 B5G 组网及关键技术。

林远卓 (1998- )，男，黑龙江哈尔滨人，北京邮电大学硕士生，主要研究方向为边缘缓存与区块链技术。

聂高峰 (1988- )，男，河南周口人，博士，北京邮电大学讲师，主要研究方向为 5G、6G 系统关键技术及移动自组织网络。

李继龙 (1976- )，男，河北邯郸人，博士，国家广播电视总局教授级高工，主要研究方向为 5G 广播、广播电视融合网、无线数字广播、信道编码和调制技术等。